

Nonlinear Storytelling: Programmierter, interaktiver Narrationsansatz für kontinuierliche Medien



Fachbereich Informatik (FB 20)
der Technischen Universität Darmstadt

genehmigte Dissertation

zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)

von

Diplom Informatiker Norbert Braun
geboren in Odernheim am Glan

Referent: Prof. Dr.-Ing. Dr. h.c. mult. José L. Encarnação
Technische Universität Darmstadt

Koreferent: Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster
Universität des Saarlandes

D17

Tag der Einreichung: 03.11.2003
Tag der mündlichen Prüfung: 09.12.2003

VORWORT

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter an den beiden INI-GraphicsNet Institutionen

GRIS Fachgebiet Graphisch Interaktive Systeme des Fachbereichs Informatik der Technischen Universität Darmstadt

ZGDV Zentrum für Graphische Datenverarbeitung Darmstadt

Mein besonderer Dank gilt den zahlreichen Personen und Institutionen, die auf sehr unterschiedliche Weise zum Gelingen der Arbeit beigetragen haben. Dazu zählen insbesondere

- Herr Prof. Dr.-Ing. Dr. h.c. mult. J. L. Encarnação für die Leitung, Durchführung und Unterstützung der Arbeit,
- Herrn Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster für sein Interesse und die Bereitschaft zur Übernahme des Koreferats,
- meine Kollegen des Fachgebiets Graphisch Interaktive Systeme - insbesondere Prof. Dr. Marc Alexa und Thomas Rieger, den Kollegen der Abteilung Animation und Bildkommunikation (A3) des Fraunhofer IGD - insbesondere Dr. Volker Luckas, Dr. Ralf Doerner, Paul Grimm und Eric Blechschmitt - sowie meinen Kollegen der Abteilung Digital Storytelling des ZGDV - insbesondere Prof. Ulrike Spierling, Prof. Manfred Gaida und Oliver Schneider - für viele produktive Anregungen und für die gute Zusammenarbeit,
- meine Diplom- und Studienarbeiter sowie meine wissenschaftlichen Hilfskräfte für Ihre Zusammenarbeit und die Ideen, die sie in ihre Arbeit eingebracht haben,
- alle Projektpartner in den Projekten EMBASSI und MAP,
- alle Kollegen im INI-GraphicsNet, die unmittelbar und mittelbar an meiner Arbeit beteiligt waren,

und meine Eltern, die mir diese nicht selbstverständliche Ausbildung ermöglicht und mich stetig unterstützt haben.

Darmstadt im Dezember 2003

Norbert Braun

KURZFASSUNG

Das Erzählen von Geschichten ist eine Form der Datenstrukturierung, welche in der natürlichen Umgebung des Menschen sehr oft zur Kommunikation zwischen Menschen eingesetzt wird. Aus diesem Grund wird das Geschichtenerzählen auch in der Informatik als Datenstrukturierung und Methode der Inhaltspräsentation genutzt. Leider gehen in diesem Zusammenhang die interaktiven Komponenten des Geschichtenerzählens zumeist verloren - der Zuhörer kann den Erzähler und damit die Erzählung nur noch unzureichend beeinflussen.

Um den oben benannten Mißstand zu überwinden wird das Konzept der interaktiven Narration von kontinuierlichen Medien vorgeschlagen. Dem Konzept liegt die Idee einer Modellierung der Geschichte basierend auf generativen (animierten) Komponenten und selektiven Komponenten (Video Clips) zugrunde. Diese Komponenten werden zur Grundlage der interaktiven Narration, dies durch Interaktionsformen zur Unterstützung der direkten Handlung des Publikums auf den Video Clips, sowie durch Interaktionsformen zur konversationalen Beeinflussung des Erzählers, und einer Einbindung der Interaktion in die Geschichtenmodellierung.

Aufbauend auf einer Diskussion der Anwendungsgebiete sowie der grundlegenden Mechanismen des interaktiven Geschichtenerzählens werden in der direkten Handlung des Publikums und in der konversationalen Beeinflussung des Erzählers die wichtigsten Komponenten des interaktiven Geschichtenerzählens erkannt. Mit diesem Wissen stellt diese Arbeit ein kombiniertes Konzept des Geschichtenerzählens sowie ein kombiniertes Konzept der Interaktion auf den Geschichten auf. Das in dieser Arbeit erstellte Konzept wird über eine Nonlinear Storytelling API für die Anwendungsprogrammierung zugänglich gemacht und anhand von verschiedenen Projekten getestet und bewertet. Die Bewertung erfolgt dabei über einen systematisch aufgestellten Anforderungskatalog, mit Hilfe dessen die grundlegenden Annahmen bezüglich der nichtlinearen Narration mit den Ergebnissen abgeglichen werden können.

INHALTSVERZEICHNIS

1. <i>Einleitung</i>	1
1.1 Anwendungen und Motivation	3
1.2 Problemstellung und Zielsetzung	6
1.3 Lösungsansatz	8
1.4 Überblick	9
2. <i>Einsatzgebiete und Anwendungsszenarien</i>	11
2.1 Avatare	13
2.1.1 Avatare als Benutzerrepräsentation	14
2.1.2 Virtuelle Personen	18
2.1.3 Avatare als Assistenten	18
2.2 Assistenz-/Agentensysteme	20
2.2.1 Beispiele einfache Assistenzsysteme	21
2.2.2 Komplexe Assistenzsysteme	23
2.3 Interaktive Narration und Digitales Video	24
2.4 Zusammenfassung	27
3. <i>Nichtlineare Storytelling Systeme</i>	29
3.1 Lineares Storytelling	29
3.1.1 Grundzüge einer Geschichte	29
3.1.2 Spannung mittels Filmtechniken	35
3.1.3 Zusammenfassung	36
3.2 Nichtlineares Storytelling	36
3.2.1 Grundlagen	37
3.2.2 Automatisierte Geschichtsnarration	43
3.2.2.1 Geschichtenbeschreibung auf Szenen-Ebene	44

3.2.2.2	Geschichtenbeschreibung auf Story-Ebene	44
3.2.2.3	Geschichtenbeschreibung auf Welt-Ebene	45
3.2.2.4	Kombinierte Geschichtenbeschreibung	46
3.2.2.5	Zusammenfassung	46
3.2.3	Funktion von interaktiv-erzählerischen Elementen	46
3.2.4	Literaturtheoretische Analyse	48
3.2.4.1	Literarische Kategorie	48
3.2.4.2	Narrative Elemente	49
3.2.4.3	Dramatische Klassifikation	49
3.2.5	Zusammenfassung	50
3.3	Aktuelle Ansätze und Konzepte	51
3.3.1	Digital Video-basierte Ansätze	51
3.3.1.1	Generische Ansätze	51
3.3.1.2	Hypercafe	55
3.3.2	VR-basierte Ansätze	57
3.3.2.1	Oz	57
3.3.2.2	Virtual Theatre and IMPROV	59
3.3.2.3	Codename: Extinct	60
3.3.3	Zusammenfassung	61
3.4	Anforderungskatalog für nichtlineares Video	61
3.5	Zusammenfassung	65
4.	Interaktion für kontinuierliche Medien	67
4.1	Konversationale Interaktion	69
4.1.1	KI und Sprachtheorie, Bottom Up Ansatz	71
4.1.2	Konversation, Top Down Ansatz	72
4.1.3	Zusammenfassung	81
4.2	Akustische Informationsdarstellung	82
4.2.1	Akustische Erweiterungen von GUI	82
4.2.2	Akustisches Browsen von Audiodaten	83
4.2.3	Akustische Annotation von Audiodaten	84
4.2.4	Akustische UI	85
4.2.5	Zusammenfassung	86

4.3	Graphische Interaktion auf kontinuierlichen Medien	87
4.3.1	Animation und VR	87
4.3.2	Video	89
4.3.2.1	Synchrone Interaktion auf Video	89
4.3.2.2	Asynchrone Interaktion auf Video	94
4.3.3	Zusammenfassung	95
4.4	Autorensysteme für kontinuierliche Medien	96
4.5	Zusammenfassung	98
5.	<i>Interaktive Narration</i>	101
5.1	Applikationsspezifisches Konzept Video	103
5.2	Benutzerinteraktion	104
5.2.1	(Video-)Applikationsspezifischer Direkt Manipulativer Zugang	104
5.2.2	Konversationaler Zugang	106
5.2.2.1	Symbolische Modellierung	106
5.2.2.2	Charakter	114
5.3	Nichtlineare Erzählstruktur	115
5.3.1	Morphologische Funktionen	117
5.3.2	Szenen	118
5.3.3	Narrativer Einsatz von Konversation	119
5.4	Zusammenfassung	122
6.	<i>Systemvalidierung</i>	125
6.1	EMBASSI	126
6.1.1	Dialogmanagement	128
6.1.2	Konversationaler Charakter	130
6.1.3	Konzeptumsetzung	133
6.2	MAP	135
6.2.1	Das MAP System	136
6.2.2	MAP Agentensystem	137
6.2.3	Der User Interface Agent	138
6.2.3.1	Agentenschnittstelle	142
6.2.3.2	Dialoginterpret	143

6.2.3.3	Konversationale Modellierung	144
6.2.4	Konzeptumsetzung	145
6.3	DIVA	146
6.3.1	Einleitung	146
6.3.2	Story & Videoservice	147
6.3.2.1	Umsetzung von Hyperlinks	148
6.3.2.2	Nichtlineare Story Narration	150
6.3.3	Konversationsaspekt	151
6.3.3.1	Charakter	152
6.3.3.2	Verhaltensentscheider	153
6.3.4	Informationshaltung & Service	154
6.3.5	Konzeptumsetzung	156
6.4	Zusammenfassung	157
7.	Anwendungen und Resultate	161
7.1	EMBASSI Anwendungen	161
7.1.1	Szenario	161
7.1.2	Anwendungstest	163
7.1.2.1	Methodik und Durchführung	164
7.1.2.2	Ergebnisse	167
7.1.3	Diskussion	169
7.2	MAP Anwendungen	170
7.2.1	Voraussetzung	171
7.2.2	Usability Beurteilung	176
7.2.3	Anwendungstest	179
7.3	DIVA Evaluation	182
7.3.1	Studien	182
7.3.2	Szenario	186
7.4	Resultate und Diskussion	188
8.	Zusammenfassung und Ausblick	191

<i>Anhang</i>	197
<i>Literaturverzeichnis</i>	199
<i>Tabellenverzeichnis</i>	216
<i>Abbildungsverzeichnis</i>	217
<i>A. Anhang</i>	223
A.1 Propp, Morphologische Funktionen	223
A.2 Explizite, temporale, intramediale Videoannotation	225
A.3 Suspense durch Filmtechniken	226
<i>B. Schriftenverzeichnis</i>	229
B.1 Eigene Publikationen	229
B.2 Betreute Diplomarbeiten	233
<i>C. Lebenslauf</i>	235

1. EINLEITUNG

Das höchste Ziel des Geschichtenerzählers ist das faszinierte Publikum, das auch nach Stunden nicht genug bekommen kann von seinen interessanten und oftmals erstaunlichen Geschichten, das nachfragt und mehr verlangt und noch am nächsten Tag begeistert mit Freunden und Bekannten über die Geschichte diskutiert. Aber warum gibt es überhaupt einen Geschichtenerzähler? Die Informationen sind doch auch ohne eine Geschichte vorhanden! Die Antwort liegt in einer Erkenntnis über die Welt, in der wir leben - sie befindet sich in einem ständigen Wandel. Aus diesem Grund hat sich seit Urzeiten Folgendes nicht geändert: Informationen stürzen massiv auf die Menschen ein. Menschen haben sich Informationen mittels Geschichten strukturiert und weitergegeben, schon in sehr früher Zeit abends am Lagerfeuer, im Mittelalter mit der Laute als Minnesänger (siehe Abbildung 1.1) oder in der Moderne mittels des Fernsehers. Die Art der Informationen blieb über die Menschheitsgeschichte im Wesentlichen gleich. Von Handlungsanweisungen für das tägliche Überleben, wie zum Beispiel der Organisation der Jagd oder des innerbetrieblichen Informationssystems, bis zur prosaischen Beschreibung einer Beziehung zwischen Menschen, wie zum Beispiel dem Roman *Romeo und Julia* [Hyl93] oder dem Spielfilm und Computerspiel *Blade Runner* [Inc98], siehe Abbildung 1.2 und Abbildung 1.3, ist alles so, wie es seit Jahrtausenden bekannt ist.



Fig. 1.1: Mittelalterlicher Geschichtenerzähler (Walter von der Vogelweide) vor Publikum [Wim98]

Ist das wirklich so? Tatsächlich ist der Zuhörer von Geschichten in der heutigen Zeit zu einem reinen Konsumenten geworden - sei es als Konsument von Büchern oder als Kon-



Fig. 1.2: Filmplakat Blade Runner

summent von Fernsehen und Film - ohne Einfluss auf den Verlauf der Geschichten. Diesen Zustand zu ändern, den Konsumenten linearer Geschichten wieder zum beeinflussenden Teil der Geschichten zu machen - und dadurch eine Nichtlinearität der Geschichten und Spontanität der Erzählweisen zu erreichen - ist das Anliegen dieser Dissertation.

Die Begrifflichkeit der linearen bzw. nichtlinearen Geschichte werden in Kapitel 3 definiert und diskutiert, an dieser Stelle soll eine Kurzdefinition der Begriffe vorgenommen werden:

- Als lineare Geschichte wird in dieser Arbeit eine Geschichte bezeichnet, welche ohne Einflussmöglichkeit von außen durch den Erzähler der Geschichte immer in derselben Weise dargeboten wird. Eine Navigation innerhalb der Geschichte - das Springen innerhalb der Geschichte zu einer anderen inhaltlichen Stelle - gilt dabei nicht als Brechen des linearen Schemas. Unzweifelhaftes Beispiel einer linearen Darstellung ist die in einem Buch niedergeschriebene Geschichte, auch wenn die Kapitel in verschiedener Reihenfolge gelesen werden können.
- Als nichtlineare Darstellung wird eine Geschichte bezeichnet, welche durch Einflussnahme von Außen - zum Beispiel durch das Publikum - in der Darbietung durch den Erzähler verändert wird. In diesem Sinn gilt das Manipulieren, Weglassen oder Hinzufügen von Inhalten, das Ändern der Sichtweise auf die Geschichte als nichtlinearer



Fig. 1.3: Online Computergame Blade Runner [Inc98]

Faktor.

Im folgenden Abschnitt wird der Trend zu nichtlinearen, interaktiven¹ Erzählweisen weiter motiviert, anschließend erfolgt ein Überblick der Arbeit.

1.1 Anwendungen und Motivation

Die Gesellschaft im 21. Jahrhundert wandelt sich von der Industriegesellschaft zur Informationsgesellschaft. Diese Wandlung begann mit der Einführung von Massenkommunikationsmitteln wie Kino, führt über das Fernsehen, welches als Teil des Lebens längst akzeptiert ist, zu einem zweiten Höhepunkt und befindet sich zur Zeit auf einem dritten Höhepunkt durch die zunehmende Verbreitung des Internet und die dadurch steigende Relevanz des Computers als Massenkommunikationsmittel.

Der Computer, einst ein Hilfsmittel für Einzelpersonen und kleine, professionelle Gruppen, wird durch das Internet zur Informationsplattform für eine breite Gruppe von Anwendern. Dies wird gefördert von immer günstigeren Internet-Providern sowie immer besseren Internetverbindungen durch höhere Netzbandbreiten auf der technischen Seite, aber auch von einer Einbeziehung des Mediums² Computer/Internet in die klassischen Massenkommunikationsmittel wie zum Beispiel Zeitungen (z.B. TV-Zeitungen mit Internet-Angeboten [TVS]) und Fernsehsendern mit Web-Angeboten [Pro], welche das Internet als zusätzliche, unterstützende Informationsplattform nutzen.

Tatsächlich werden die Medienangebote nicht nur durch Computer/Internet erweitert. Die vom Computer her gewohnte Möglichkeit der Interaktion mit den Medien hat Konsequenzen

¹ Nach Bahrddt [Bah90] wird als Interaktion ein wechselseitiges, soziales Handeln von zwei oder mehr Personen bezeichnet, siehe Kapitel 4.

² Der Begriff Medium ist in diesem Zusammenhang im Sinne von Präsentationsmedien gewählt. Eine Definition der Begrifflichkeit Medium ist in Kapitel 4 zu finden.

für das traditionelle Medium Fernsehen. Durch Interaktives Fernsehen wird ein Teil der vom Computer gewohnten Interaktivität auf das Fernsehen übertragen. Mittels Set-Top-Boxen oder DVD werden Pay per View, Spartenprogramme, Indizierung, freie Kamerawahl und Zusatzinformationen für den Zuschauer zur Realität. Neben dieser Entwicklung vom Computer hin zum Fernseher wird durch Filmsequenzen in Videospielen auch eine Entwicklung vom Fernsehen/Kino zum Computer sichtbar. So werden Medienpräsentationen auf dem Computer zunehmend von Video³-Sequenzen ergänzt und sogar bestimmt. Neben diesen Anwendungen wird das Fernsehen als solches digital über das Internet übertragen - mittels *Streaming*⁴ Video als *Internet Broadcast* (Internet-Angebote für Radio und TV, die eine eigene Form des E-Commerce darstellen). Vom Internet Broadcast verlangt der Benutzer mittlerweile die Einflussmöglichkeiten, die er von Multimedia-Präsentationen her gewohnt ist, d.h. ein Grundmaß an Interaktivität, dass über die Funktionalität eines Videorecorders hinausgeht. An dieser Stelle ist der Einsatz von nichtlinearem Storytelling⁵ durch interaktives Video gefordert. Wird diese Funktionalität geboten, so wird dies als *Interactive Internet Broadcast* bezeichnet. Im extremsten Falle wird das klassische Video vollständig durch eine interaktive virtuelle⁶ Realität ersetzt, welche einen interaktiven Spielfilm (Interactive Movie) formt. Diese Motivation führt zum Aufstellen der ersten Arbeitshypothese:

Arbeitshypothese 1: Interaktivität kann als dramaturgisches⁷ und immersions-steigerndes⁸ Mittel vorteilhaft zur Narration⁹ von kontinuierlichen¹⁰ Medien, zum Beispiel Video, eingesetzt werden.

Neben den rein geschäftlichen und technischen Aspekten dieser Internet-Revolution wird ein weiterer Aspekt der Informationsgesellschaft über die Erweiterung der Medienlandschaft um den Computer beeinflusst: Fernsehen ist Teil des sozialen Lebens, die im Fernsehen auftretenden Figuren werden als parasoziale Komponente des sozialen Lebens akzeptiert [Gle97]. Jedoch fehlen dem klassischen Fernsehen die notwendigen schnellen Interaktionsmöglichkeiten durch den Zuschauer. Die parasoziale Akzeptanz überträgt sich in erweitertem Maße auf künstliche Figuren in virtuellen Welten: Aufgrund deren Fähigkeit, auf den Benutzer zu reagieren, werden diese Figuren als soziale Wesen erlebt. Dadurch erfährt die Schnittstelle

³ Ein Video besteht aus zwei Medien - dem bewegten Bild (eine Sequenz von Bildern, welche dem menschlichen Auge als kontinuierlich erscheinen) und dem Ton. Diese Medien werden parallel abgespielt und als eine mediale Einheit angesehen.

⁴ Der Begriff Streaming bezeichnet das kontinuierliche Versenden von Daten. Im Bereich Video bezeichnet er im Allgemeinen eine Möglichkeit, ein Video zu präsentieren ohne vorher das gesamte Video zur Verfügung zu haben - die einzelnen Videoteile werden im Augenblick ihrer Verfügbarkeit präsentiert.

⁵ Storytelling: Der englische Ausdruck für Geschichtenerzählen. Der Begriff Story entspricht dem - leider mehrdeutigen deutschen Begriff - Geschichte. Geschichte ist im Zusammenhang dieser Arbeit immer nach Kapitel 3.1.1 definiert. Wird Geschichte als Historie oder Historik angesehen, so ist dies explizit vermerkt.

⁶ Der Begriff Virtuell bezeichnet die Eigenschaft, etwas im Wesen oder in der Wirkung, jedoch nicht Wirklich zu sein.

⁷ Der Begriff des Dramas und der Dramaturgie wird in Kapitel 3.1.1 spezifiziert. Ein Drama ist eine spezielle Form der Geschichte, eine Dramaturgie bezeichnet eine gewisse Handlungsabfolge in einer Geschichte.

⁸ Immersion wird in diesem Zusammenhang gedeutet als das eintauchen/aufgehen in eine Thematik/Geschichte.

⁹ Narration bezeichnet das Erzählen (im Sinne von vermitteln) einer Geschichte, siehe Kapitel 3.1.1.

¹⁰ Kontinuierlich bedeutet nicht unterteilbar in kleinste Zeiteinheiten. Im Gegensatz dazu steht der Begriff Diskret: Er bezeichnet die Möglichkeit der zeitlichen Einteilung in kleinste Einheiten.

zwischen Mensch und Computer eine Aufwertung ihrer sozialen Bedeutung - Benutzern wird das Verstehen durch Ansprechen ihrer Gefühlsebene erleichtert. Neben dem Ruf nach neuen Präsentations- und Interaktionsmethoden für Fernsehen und Video, sowohl im Heim- als auch im Office-Bereich, führt dies unter anderem zur Schaffung von künstlichen Stars oder Moderatoren, z.B. der japanischen Kunstfigur Kioto Date [Dat].

Der obere Abschnitt geht konform mit der aktuellen Tendenz der klassischen Informatik, weg von der standardisierten Benutzerinteraktion mit den Geräten Screen, Tastatur und Mouse und dem WIMP-Paradigma (Windows, Icons, Menus, Point&Click) hin zu medienbezogener, menschähnlicher Interaktion [Bux93] sowie einer Erweiterung der genutzten Menge der Medien (neben den bisher schwerpunkthaft genutzten Medien Text und Bild) um die Medien Audio, Video, Animation¹¹, Virtual Reality¹² (VR) und Augmented Reality¹³ (AR). Diese Erweiterung um kontinuierliche Medientypen führt zur Forderung nach erzählerischen Interfaces mit einer benutzerspezifischen Präsentation von Daten. Für den Autor von Präsentationen wird eine *Programmierschnittstelle für Jedermann* zur Komposition der Medien [ES99] gefordert.

Hier führen die beiden Tendenzen (Annäherung zwischen klassischen Medien und Informatik) ein weiteres Mal zusammen, nicht auf einer technischen, sondern intellektuellen Ebene. Narration soll durch Erzählerische Interfaces vor allem die Immersion in den Kontext der Informationsvermittlung steigern. Information können durch die Immersion des Benutzers besser verarbeitet und damit erlebt oder erinnert werden. Durch Interaktion wird Langeweile vermieden sowohl im Entertainment-Bereich als auch in klassischen, ernsthaften Informatik-Anwendungen, vergleiche [Enc99a].

Die erzählerischen Interfaces können über die bisherige Nutzung des Computers als Werkzeug für Spezialisten nicht realisiert werden. Vielmehr ändert sich die Metapher der Mensch-Maschine Interaktion hin zum Assistenten des Benutzers mit einem situationsbedingten, delegativen Dialog-Paradigma, vergleiche [WRB01]. Diese Metapher setzt den Schwerpunkt auf eine natürliche Weise der Kommunikation, also eher Mensch-zentriert mit Focus auf die angebotenen Medien und Informationen, als auf den Aspekt der Systembedienung. Idealerweise findet die Kommunikation mit der Maschine als menschähnliche Konversation¹⁴ statt, vgl. Braun [Bra03a].

Im gesellschaftlichen Alltag ist der Computer als Gesprächspartner bereits akzeptiert - er existiert zwar nicht in der Realität, aber in der Vorstellungswelt: Seit Generationen werden Computer in Film und Fernsehen als allwissende, intelligente und eloquente Gesprächspartner gezeigt. So wird zum Beispiel in dem 1968 entstandenen Stanley Kubrick Film *2001: A Space Odyssey* [Cla68] ein Computer als zentrales Element des Films gezeigt - als einer der (Haupt-) Protagonisten. Ebenso ist in Gene Rodenberrys Fernsehserie *Star Trek*

¹¹ Als Animation wird eine, aus computergenerierten Bildern bestehende, Bild-Sequenz bezeichnet.

¹² Als Virtual Reality wird die Simulation einer realen oder imaginären Umgebung bezeichnet, welche durch einen Benutzer in den drei Dimensionen Länge, Höhe und Breite erfahren werden kann. Oft wird zusätzlich eine interaktive Möglichkeit der Beeinflussung der Umgebung durch den Benutzer gefordert.

¹³ Als Augmented Reality wird eine reale Umgebung bezeichnet, in welche eine künstliche Umgebung (Virtual Reality) nahtlos integriert ist. Ein Benutzer kann beide Umgebungen überlagert wahrnehmen.

¹⁴ Der Begriff der Konversation ist in Kapitel 4.1 definiert.

[Ale94], entstanden zwischen 1966 und 1969, der Computer als konversational anzusprechendes, allwissendes Hilfsmittel präsentiert. Dies steigerte die Erwartungen der Benutzer von Assistenzsystemen weit über das Maß, welches zu dieser Zeit (und auch heute) technisch realisierbar war.

Der obere Abschnitt lässt das Aufstellen einer weiteren Arbeitshypothese zu:

Arbeitshypothese 2: Konversation kann als Metapher¹⁵ der Kommunikation mit Computersystemen vorteilhaft zur Assistenz und Delegation durch den Benutzer eingesetzt werden.

Durch die Zentralisierung des Menschen in diesem Ansatz wird eine neue Qualität des Designs von Schnittstellen zwischen Mensch und Maschine angestrebt. Dieses Design muss an die Medien, mit denen Informationen interaktiv via Geschichten vermittelt werden, angepasst sein. Es erfordert eine Zusammenarbeit verschiedener Experten auf den Gebieten Storytelling (z.B. Film, Fernsehen), Design und Informatik [BGS99], um das Ziel der erzählerischen Interfaces und mehr noch, der direkten Kommunikation des Publikums mit den präsentierten Inhalten, zu erreichen.

1.2 Problemstellung und Zielsetzung

Wie im vorherigen Abschnitt beschrieben, existiert eine Notwendigkeit, die lineare Struktur der Narration, z.B. des Mediums Video, aufzubrechen und mittels interaktiver Komponenten zu einer nichtlinearen Narration von Inhalten zu gelangen. Die bisher bekannten Ansätze berücksichtigten jedoch ausschließlich direkt manipulative Ansätze, welche im wesentlichen auf hypermedialen¹⁶ Strukturen für Stillbilder beruhen. Durch das fehlende interaktive Element für Audio wird ein Großteil der Informationen für die direkte Manipulation durch den Benutzer unzugänglich. Bisher fehlt zudem ein Assistenz-Ansatz, der mittels konversationaler Interaktion eine weitere, natürliche Art des nichtlinearen Geschichtenerzählens bietet. Im Folgenden wird der direkt manipulative wie der konversationale Ansatz umrissen und die Anforderungen an nichtlineares Geschichtenerzählen mit Video aufgelistet.

Die Kunst des nichtlinearen Geschichtenerzählens wird wesentlich durch die Möglichkeiten des Publikums, die Geschichte interaktiv zu beeinflussen, geprägt. Die klassische Art der Interaktion mit einem Erzähler ist die Konversation - erst in zweiter Linie erfolgt die direkte, eigene Handlung des Publikums. Je nach Struktur der Geschichte wird eine bestimmte, spezielle Art der Interaktionen mit den Medien, welche die Geschichte erzählen, notwendig. Jedes dieser Medien besitzt eine eigene, spezifische Charakteristik, die die Art der direkt manipulativen Interaktion auf dem Medium entscheidend beeinflusst. Diese Charakteristika, wie zum Beispiel zeitliche Dauer, spatiale Ausprägung, müssen bei der Entwicklung eines Interaktionsansatzes auf dem entsprechenden Medium besonders berücksichtigt werden - der Interaktionsansatz muss somit Medien-spezifisch sein. Eine Hyperlink-Annotation

¹⁵ Der Begriff Metapher geht in Wort und Bedeutung auf den griechischen Philosophen Aristoteles zurück. So gilt seit zweitausend Jahren die Metapher für die (bewusste) Übertragung einer Benennung, die eigentlich etwas Anderes bedeutet, sei es die Übertragung vom weiteren Begriff auf den engeren, oder vom engeren Begriff auf den weiteren [Mau01].

¹⁶ Der Begriff Hypermedia ist in Kapitel 4 definiert.

in zeitabhängigen Medien ist zum Beispiel nur für einen gewissen Zeitraum aktiv. Das Medium Video kann zeitweise durch ein Polygon graphisch annotiert werden [CTCL96] oder durch einen Sound akustisch annotiert werden [BD98b]. Diese zeitabhängige Form der Interaktionsofferte wird synchrone Interaktion genannt, da sie zeitlich synchron zum Video verläuft.

Neben der synchronen Interaktion sollte dem Benutzer auch eine nicht-zeitgebundene, asynchrone Interaktionsmöglichkeit offen stehen. Diese Interaktionsmöglichkeit ist als Medienunabhängig einzustufen. Durch diese Unabhängigkeit ergibt sich die Möglichkeit, spezielle, menschähnliche Kommunikationsmethodiken mit sozialen und emotionalen Aspekten, wie zum Beispiel die Konversation, zu berücksichtigen. Ein konversationales Interaktionsszenario kann zum Beispiel auf einem Assistenzsystem¹⁷ mit einer Avatar-Repräsentanz¹⁸ aufbauen [Mat97, SP95]. Dieser Avatar wird pro-aktiv durch die Informationen des Videos angetrieben und bildet auf diese Weise eine sozial-emotionale, konversationale Schnittstelle der nichtlinearen Narration. Mittels geeigneter konversationaler Interaktion kann eine sehr freie, benutzerspezifische Narration erreicht werden.

Diese Arbeit fordert die Kombination dieser beiden Fälle, damit der Benutzer sowohl als Konsument als auch als Beeinflusser einer Geschichte tätig werden kann. Aus diesem Grund werden synchrone und asynchrone Interaktionsmöglichkeit im Video vereint:

- Die synchrone Interaktionsmöglichkeit führt zu einer direkt manipulativen Interaktionsmöglichkeit, die speziell auf das Medium Video optimiert werden soll.
- Die asynchrone Interaktionsmöglichkeit führt zu einer konversationalen Interaktionsmöglichkeit auf Video, die den menschlich-sozialen und emotionalen Aspekt der Kommunikation berücksichtigen soll.

Nichtlineare Videos werden bisher nur in äußerst begrenztem Umfang hergestellt, da es sowohl an Authoring-, Service- und Präsentationstools als auch an einer wissenschaftlichen und künstlerischen Basis für solche Videos fehlt. Die Nutzung von linearen Videos als Ausgangsmaterial ist z.Z. die Regel, obwohl diese nicht für den interaktiven Einsatz geplant sind. Die Zusatzinformationen, welche für einen erweiterten Einsatz von Video für nichtlineare Erzählungen benötigt werden, sind bisher nicht definiert worden, auch gibt es keine ausgereiften Konzepte, um auf sie zuzugreifen. So werden die heute zugreifbaren nichtlinearen Geschichten mit den Methodiken des klassischen Films erstellt, was die Qualität dieser Geschichten reduziert. Diese Arbeit fordert ein Konzept zur Beschreibung interaktiver nichtlinearer Geschichten durch Video, welches die Narration durch den Autor unterstützt und gleichzeitig eine benutzerspezifische Narration mit freier Dramaturgie, echten Spannungsbögen und entsprechender Benutzeremotion ermöglicht.

Um den Aufwand bei der Erstellung von nichtlinearen Videopräsentationen zu reduzieren, wird ein Application Programmer's Interface (API) auf einem hohen Abstraktionslevel mit Focus auf Narration und Interaktion zwischen Videopräsentation und Benutzer benötigt.

¹⁷ Siehe Definition Kapitel 2.2.

¹⁸ Avatar: Personifizierter Repräsentant, siehe Definition Kapitel 2.1.

Diese sollte den gestalterischen und inhaltlichen Anforderungen verschiedener Nutzungsmodelle von interaktivem, nichtlinearem Video gerecht werden. Die Animation von konversationalen Elementen (zumeist ein sehr aufwendiges Unterfangen) sollte auf ein Minimum reduziert werden. Das in dieser Arbeit geforderte System soll die entsprechenden Services für synchrone und asynchrone Interaktion in einer generischen Form bieten und damit die Konzentration des Autors auf die Geschichte, statt auf die Technik, gewährleisten.

An dieser Stelle soll die Zielsetzung dieser Arbeit noch einmal aufgelistet werden:

- Konzepte zur Beschreibung interaktiver nichtlinearer Geschichten durch Video.
- Verknüpfung von synchroner, direkt manipulativer und asynchroner, konversationaler Interaktionsmöglichkeit, um dem Benutzer einen zweckdienlich-intuitiven und trotzdem menschähnlich-sozialen Einfluss auf die Präsentation des nichtlinearen Videos zu ermöglichen.
- Entsprechende Service- und Präsentationssysteme in einer generischen Form, die den gestalterischen und inhaltlichen Anforderungen verschiedener Nutzungsmodelle von interaktivem nichtlinearem Video gerecht werden und über API für den Anwendungsprogrammierer nutzbar sind.

1.3 Lösungsansatz

Der in dieser Arbeit vorgestellte Lösungsansatz sieht eine Trennung von Story-gebender und Konversations-modellierender bzw. Direkt Manipulativer Methodik vor, vgl. Abbildung 1.4.

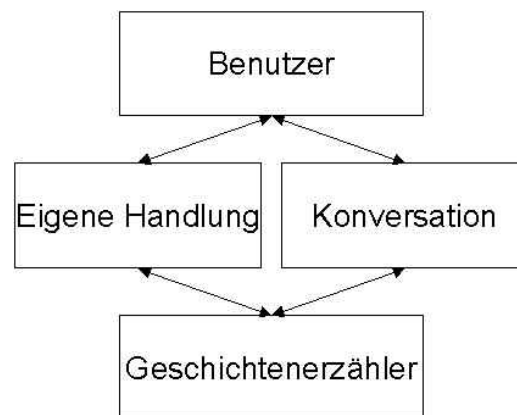


Fig. 1.4: Trennung der Methoden

Durch diese Trennung kann auf die jeweilige Problematik der einzelnen Gebiete gesondert eingegangen werden und die Lösungen der Gebiete - auch separat - auf weitere Anwendungen übertragen werden.

So wird bezüglich Storytelling ein morphologischer Ansatz des interaktiven Geschichtenerzählens empfohlen, der trotz seiner Ausrichtung von der Ausgangssituation bis zum Finale

über spannungssteigernde Momente eine interaktive und spontane Erzählweise zulässt. Die direkte Handlung des Publikums, d.h. die direkt manipulative Interaktion, wird medienspezifisch gelöst - bezüglich Videointeraktion wird ein Ansatz der temporalen, expliziten und intramedialen Annotation empfohlen. Die eigentliche, natürliche Interaktionsform mit Geschichtenerzählern, die Konversation, wird separat und explizit modelliert, da eine Trennung von Kontext und Konversationsverhalten dem menschlichen Vorbild am nächsten kommt und einen Design-bezogenen, revisionsfreundlichen Ansatz erlaubt. Durch die Berücksichtigung dieser drei Aspekte kann ein generisches System zum interaktiven Erzählen von Geschichten entworfen werden, das einerseits keiner speziellen Ästhetik verpflichtet ist, andererseits aber die grundlegenden Funktionen und API zur Ausformulierung einer speziellen Ästhetik besitzt, die tatsächlich funktioniert¹⁹.

1.4 Überblick

Dieser Abschnitt enthält eine thematische und inhaltliche Kurzfassung der vorliegenden Arbeit. Die Arbeit hat zum Ziel, einen Ansatz zur interaktiven Narration von kontinuierlichen Medien am Beispiel digitales Video zu bieten. Zu diesem Zweck werden die Anforderungen nichtlinearen Geschichtenerzählens mit interaktivem Video systematisch analysiert. Die Interaktionsformen auf kontinuierlichen Medien werden insbesondere auf die Anforderungen der synchronen und asynchronen Interaktion auf Video geprüft. Das Konzept der akustischen Hyperlinks wird erarbeitet, dieses wird im Hinblick auf die Kombination mit graphischer Annotation von Video und intermedialen Interaktionsformen kritisch bewertet und in den Zusammenhang mit graphischer (point & click) und konversationaler Interaktion gestellt. Ein Konzept zur symbolischen Modellierung von Konversation wird erstellt, desweiteren ein Konzept zur symbolischen, morphologischen Modellierung von interaktiven Geschichten notiert, ein Ansatz zur intramedialen, expliziten und temporalen Annotation von Video eingeführt und die verschiedenen Konzepte in einem Gesamtansatz zusammengeführt. Entsprechende Systemarchitekturen werden diskutiert und auf die Erfüllung der Anforderungen hin überprüft.

Der Aufbau der Arbeit: Im aktuellen Kapitel wird die Arbeit motiviert und die Fragestellungen der Arbeit sowie die Lösungsidee skizziert. Zur Identifikation der durch synchrone und asynchrone Videointeraktion potentiell lösbaren Problemstellungen in dem Anwendungsgebiet nichtlineare Geschichten wird eine Anwendungsanalyse in Kapitel 2 durchgeführt. Diese Anwendungsanalyse zeigt das Fehlen eines Konzeptes für Nonlinear Storytelling. Die formulierten Problemstellungen der Einleitung bezüglich nichtlinearem Geschichtenerzählens mit Hilfe von synchron- und asynchron interaktivem Video werden aus diesem Grund in Kapitel 3 weiter präzisiert, u.a. durch eine State of the Art Analyse zum Thema nichtlineares, interaktives Geschichtenerzählen. Danach wird in Kapitel 4 eine weitere State of the Art Analyse zur Thematik interaktives Video im Allgemeinen und im Speziellen auf die Nutzung von hypermedialen und konversationalen Interaktionsmethaphern durchgeführt. Zur Thematik

¹⁹ Der Ausdruck wird hier im künstlerischen Sinne des Funktionierens von Geschichten durch eine Erzeugung von Spannung und Immersion gebraucht.

akustische Userinterfaces wird das Konzept der akustischen Annotation entwickelt. Die akustische Annotation wird anschließend in die Konzepte synchrone Videointeraktion und asynchrone Videointeraktion auf Graphik und Akustik des Videos eingeordnet. In Kapitel 5 werden die in den vorherigen Kapiteln aufgestellten Forderungen und Schlußfolgerungen in einem Konzept zur interaktiven Narration von kontinuierlichen Medien am Beispiel Video vereint. Im Kapitel 6 werden Systeme und deren Architektur beschrieben, welche die Forderungen an nichtlineares Geschichtenerzählen mit interaktivem Video erfüllen und auf der Nutzung von graphisch-akustisch-synchroner und -asynchroner, konversationaler Interaktion basieren. Diese Systeme dienen der Validierung des Konzeptes.

Unter Verwendung der Systeme werden in Kapitel 7 Szenarien aufgestellt und die Auswertung dieser Szenarien beschrieben. Die in den Szenarien erzielten Resultate werden vor dem Hintergrund der in Kapitel 3 und 4 ermittelten Anforderungen diskutiert. Schließlich wird die Arbeit zusammenfassend in Kapitel 8 betrachtet und mit einem Ausblick auf thematisch weiterführende Forschungsaktivitäten und potentielle Forschungsvorhaben abgerundet.

In den Anhängen sind Literaturverzeichnis, Bild- und Tabellenverzeichnis zu finden, eine Definition des Nonlinear Storytelling API findet sich in [Bra01c].

2. EINSATZGEBIETE UND ANWENDUNGSSZENARIEN

Die Kunst des Geschichtenerzählens wird zur Zeit nur in wenigen Systemen explizit zum Nutzen des Benutzers eingesetzt - narrative Ansätze in kommerziellen Anwendungen finden sich äußerst spärlich. Interaktivität wird dabei oftmals nicht angestrebt - so wird die Narration von Geschichten oft für das Notieren von Geschichten mittels digitaler Medien belegt und durch den Ausdruck *Digital Storytelling* bezeichnet¹, es besteht jedoch kein Zusammenhang zu interaktiver Narration.

Die vorliegende Arbeit folgt einer Definition von Encarnação, vgl. [Enc97], aus dem Bereich der Computer Graphik - diese ist sehr eng mit dem Begriff der interaktiven Narration verknüpft: *Digital Storytelling bezeichnet die integrierte, multimediale Erfahrung des Benutzers, welche durch eine Orchestrierung aller multimodalen² und multimedialen Möglichkeiten erzeugt wird.* In der entsprechenden Literatur ist eine Fokussierung auf die zu erzählende Geschichte explizit mit einbezogen, vgl. [Enc99b].

Derzeit muss ein Benutzer oftmals vom Computer dargebotene Informationen, die nicht in Form einer Geschichte strukturiert sind, selbst strukturieren³. Diese Strukturierungsarbeit kann einem Benutzer wenigstens erleichtert, in vielen Fällen sogar gänzlich abgenommen werden. Die generelle Ausrichtung dieser Systeme geht somit in Richtung der Assistenz des Benutzers: Assistenz beim Suchen und Strukturieren von Informationen, aber auch Assistenz bei der Bedienung von Geräten durch anschauliche Darstellung dieser Bedienung oder beim Umgang mit Medien durch medienzentrierte Interaktion.

Assistenzsysteme müssen in besonderem Maße auf die Bedürfnisse von Benutzern zugeschnitten sein. Im Allgemeinen wird daher verlangt, dass der Benutzer die notwendige Kommunikation mit einem Assistenzsystem in einer natürlichen, der menschlichen Konversation sehr nahe kommenden Art führen kann, vgl. [EK00]. Dies bedingt unter Anderem, dass der Benutzer einen festen Ansprechpartner zur Delegation von Aufgaben und zum Entgegennehmen von Informationen besitzt. Dieser Ansprechpartner könnte ein menschähnliches Antlitz besitzen, aber auch eine künstlerisch verfremdete, menschähnliche Sache (z.B. sprechende Blume) sein. Es existieren diverse Ansätze im Gebiet der Virtual Reality und Animation, die sich mit der Realisierung von solchen anthropomorphen Assistenten beschäftigen.

Weiterhin sollte die Interaktion mit dem Assistenzsystem auf eine intuitive, wenn möglich

¹ Vgl. zum Beispiel das Digital Storytelling Festival [Atc01].

² Der Begriff der Modalität ist in Kapitel 4 definiert.

³ Informationen sind zumeist in einer Weise strukturiert, welche der Verarbeitung durch einen Computer entgegenkommen, z.B. Relational, als Netzwerk oder Objektorientiert. Diese Strukturierung entspricht nicht der Weise, wie ein Mensch Informationen strukturiert.

natursprachliche Art oder durch Verwendung von Zeigegestik erfolgen. Tatsächlich werden mit diesem Anspruch verschiedene Paradigmen der Mensch-Maschine Interaktion bemüht - nämlich sowohl der direkt manipulative Ansatz (Zeigegestiken auf bestimmte Objekte ermöglichen die direkte Interaktion mit diesen Objekten) als auch der konversationale Ansatz (natursprachliche Interaktion führt oftmals zur Konversation zwischen Benutzer und Maschine - dies bedingt i.a. eine indirekte Manipulation von Objekten).

Wie oben beschrieben benötigt eine natürliche, flexible, effiziente und leistungsfähige Mensch-Maschine Kommunikation neben der generellen Assistenz jedoch noch einen weiteren Punkt: Die Einbettung der Assistenz in einen thematischen und strukturellen Kontext, der dem Benutzer hilft, die Informationen zu verstehen. Video⁴ Clips⁵ sind ein sehr effizientes Mittel, Informationen innerhalb einer verständnisfördernden Umgebung zu präsentieren. Der Ansatz in dieser Arbeit bezieht Video-Clips explizit mit ein. Dies hat verschiedene Vorteile:

- Natürlichkeit der Darstellung: Video bietet eine relativ natürliche Darstellungsform für verschiedenste Dinge wie zum Beispiel industrielle Prozesse, Entertainment, Pädagogik etc.. Video vermittelt den Eindruck des 'Echten' derzeit in erheblich stärkerer Form als zum Beispiel Virtual Reality.
- Ressourcen: Sehr viele Video-Clips sind bereits gedreht worden und in Archiven abgelegt. Diese Video-Clips können - trotz der darin verwendeten linearen Narration - weiter verwendet werden.
- Autorenangebot: Der Bereich Video ist bereits etabliert, die entsprechenden Produktionstechniken sehr ausgereift. Dies führt dazu, dass Autoren von Geschichten sehr oft zu Video als Medium zurückgreifen.
- Publikumsakzeptanz: Das Publikum einer Geschichte ist es in starkem Ausmaß gewohnt, Informationen mittels eines Videos zu beziehen.

Wird Video als Medium zur Präsentation genutzt, so sollte das Video nahtlos in die entsprechenden Systeme integriert sein. Dies bedeutet, dass ein interaktiver, konversationaler Zugriff auch auf Informationen in Videos notwendig ist. Video sollte in eine Narrationsstruktur integriert sein und das Erzählen von Geschichten unterstützen. Neben dem Video sollten unbedingt Techniken der Animation (z.B. Avatare) genutzt werden, um den oben beschriebenen, festen Ansprechpartner des Benutzers zu gewährleisten. Es liegt daher sehr nahe, graphisch-interaktive Interaktionstechniken und akustisch-interaktive Interaktionstechniken zur videobasierten Interaktion mit Ansätzen aus dem Bereich Virtual Reality (anthropomorphe Assistenten) zu kombinieren, um einen effizienten, hybriden Ansatz der Mensch-Medien-Interaktion zu erzielen.

Dieser Ansatz ist zur Zeit noch in keiner Anwendung realisiert - daher können die einzelnen Komponenten des Ansatzes lediglich isoliert in verschiedenen Ansätzen gefunden werden.

⁴ In der vorliegenden Arbeit wird ein Video als digitales Medium verstanden - d.h. die entsprechenden Daten liegen in digitaler Form vor.

⁵ Die Bezeichnung Clip intendiert eine gewisse temporale Beschränkung des Videos.

Die Einsatzgebiete der Technologien liegen bisher im Bereich von Web-basiertem Verkauf, also im sogenannten E-Commerce. Weitere Einsatzgebiete befinden sich im künstlerischen Bereich, bei Entertainment-Anwendungen - und im Home-Bereich. Benutzer im Home-Bereich sind im Allgemeinen für den Umgang mit Direct Manipulation-Interfaces nicht ausgebildet. Gerade im Home-Bereich ist eine Assistenz-Schnittstelle für, z. B. die Bedienung der Hausgeräte oder des interaktiven Fernsehens, eine Lösung, welche die Hürden der komplizierten Bedienung von Geräten umschifft.

Im weiteren werden einige Anwendungsbereiche für Avatare, Assistenten im Bereich E-Commerce/World Wide Web (WWW) geschildert sowie die kommerziellen Anwendungen im Bereich Digitales Video und Nichtlineare Narration diskutiert. Spezielle Forschungen auf den oben genannten Gebieten werden in Kapitel 3 und Kapitel 4 diskutiert.

2.1 Avatare

Avatare müssen nicht unweigerlich in Assistenzsystemen Verwendung finden. Ursprünglich wurden Avatare im wesentlichen zur virtuellen Darstellung von Benutzern in Multiuser-Dialogsystemen verwandt. Zum besseren Verständnis soll daher an dieser Stelle eine Definition des Begriffes *Avatar* erfolgen.

Avatar stammt aus dem Sanskrit, vgl. [Wil98]. Demgemäß brauchen Gottheiten Körper, damit sie sich auf der Erde bewegen und mit den Menschen kommunizieren können. Diese Manifestationen werden Avatare genannt.

Dieses Prinzip der Verkörperung durch Avatare – das heißt der Einsatz künstlicher Repräsentationen – ist u.a. im *Cyberspace* (eine künstlich geschaffene, Multiuser-VR-Welt) zu finden.

Die Ausprägungen von Avataren können anhand von Beispielen beschrieben werden, vgl. [Wil98]. Folgende Ausdrucksformen sind zu den Avataren zu zählen:

- 3D-Abbildungen, die vom Aussehen dem Benutzer nachempfunden sind und deren Bewegungen nur von ihm gesteuert werden.
- Darstellungen, deren generelles Verhalten der Benutzer bestimmt, deren Gestik etc. aber vom System gesteuert werden. Als Beispiel wird das *Oz*-Projekt der Carnegie Mellon University (siehe Kapitel 3.3.2.1) angeführt.
- Virtuelle Personen mit virtuellen Körpern und Live-Video-Gesichtern, wie sie beispielsweise in NTTs *InterSpace* vorkommen (zu sehen in Abbildung 2.1), sofern der Benutzer die Bewegungen steuern und mittels gesprochenem oder geschriebenem Text kommunizieren kann.
- Darstellungen, die - vom Benutzer kontrolliert - Emotionen ausdrücken können, auch wenn einige Ausdrucksmöglichkeiten vom System gesteuert werden.

- Abbildungen, die selbständig Routinen ausführen, auch wenn dabei vor Allem von Artificial Intelligence Gebrauch gemacht wird, z.B. die Zusammenstellung von Kreditkartenzahlungen, die während des Internet-Shoppings vom Benutzer getätigt wurden.

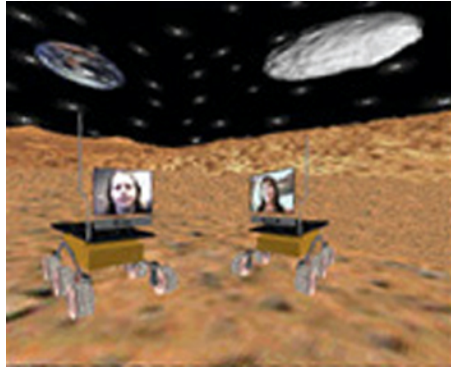


Fig. 2.1: Avatare der Softwarefirma NTT mit Live-Video-Gesichtern, Screenshot [Cor01]

Die derzeit existierenden Avatare können anhand ihrer Aufgaben in folgende Klassen eingeteilt werden:

- Avatare als Benutzerrepräsentation in virtuellen Umgebungen
- Virtuelle Persönlichkeiten
- Avatare als Assistenten

In den folgenden Kapiteln werden Anwendungsbeispiele zu den Klassen gegeben.

2.1.1 Avatare als Benutzerrepräsentation

In Anwendungen mit mehreren Benutzern, die sich nicht lokal gegenseitig sehen können, ist eine virtuelle Repräsentanz dieser Benutzer oftmals eine sehr natürliche Art, eine gewisse Vorstellung vom anderen Benutzer zu erhalten.

Gerade das World Wide Web bietet dem Benutzer nicht nur umfangreiche Informationen, sondern auch die Möglichkeit, mit anderen Usern in Kontakt zu treten. Dies erfolgt durch die Verwendung von E-mail, in Chatrooms bzw. in anderen virtuellen Räumen. Dadurch wurde der Aufbau eines eigenen Gesellschaftssystems mit entsprechenden sozialen Gruppierungen ermöglicht (vgl. [KA98]). Beispiele für diese Treffpunkte im Internet sind die künstlichen Welten "Virtual Computer" oder "Godzilla", vgl. [act00].

Der Benutzer erhält in der virtuellen Umgebung eine Repräsentation, damit er von den anderen Teilnehmern wahrgenommen werden kann. Die Art dieser Repräsentation sowie das Erscheinungsbild der Internet-Treffpunkte stellen sich sehr unterschiedlich dar. Es gibt Umgebungen, die aus Elementen wie Listen, Scrollbalken und Formularfeldern bestehen

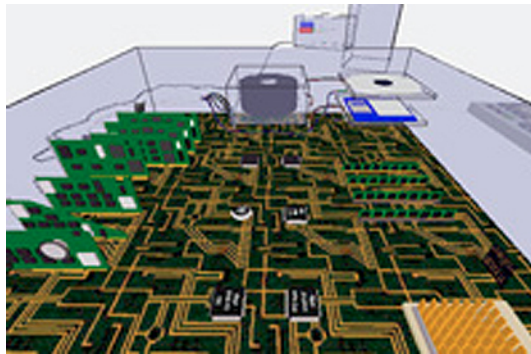


Fig. 2.2: Virtual Computer von Activeworlds [act00]

und in denen die Benutzer meist nur durch Email-Adressen oder Nicknames repräsentiert werden (siehe Abbildung 2.3).

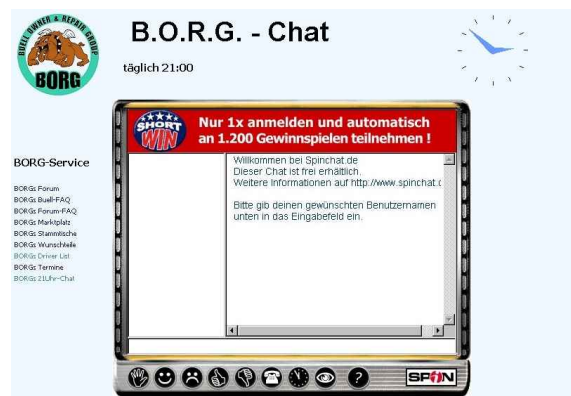


Fig. 2.3: Formularbasierter Chatroom

Inzwischen gibt es aber auch virtuelle Welten, virtuelle Städte und Chatrooms, in denen sich die Benutzer in virtuellen Landschaften oder Gebäuden etc. treffen können. Auch diese Umgebungen differieren in grafischer und qualitativer Hinsicht sehr stark. Sie umfassen einfache 2D-Welten ebenso wie nahezu realistische 3D-Darstellungen. Ein Beispiel dafür ist die in Abbildung 2.4 [cyb00b] dargestellte virtuelle Stadt "Cybertown".

Diese Welten bieten dem Benutzer die Möglichkeit der Repräsentation mittels der grafischen Verkörperung durch einen Avatar. Dieser kann sich unter der Kontrolle des Benutzers durch die virtuellen Welten bewegen und ist in dieser Form auch für andere sichtbar. Erst die Verwendung von Avataren macht es den Benutzern möglich, sich bei ihren sozialen Kontakten im Cyberspace der menschgewohnten Kommunikation mit all ihren Ausprägungen anzunähern, vgl. [Doy99]. Die Stärke dieser Annäherung hängt von der Art des eingesetzten

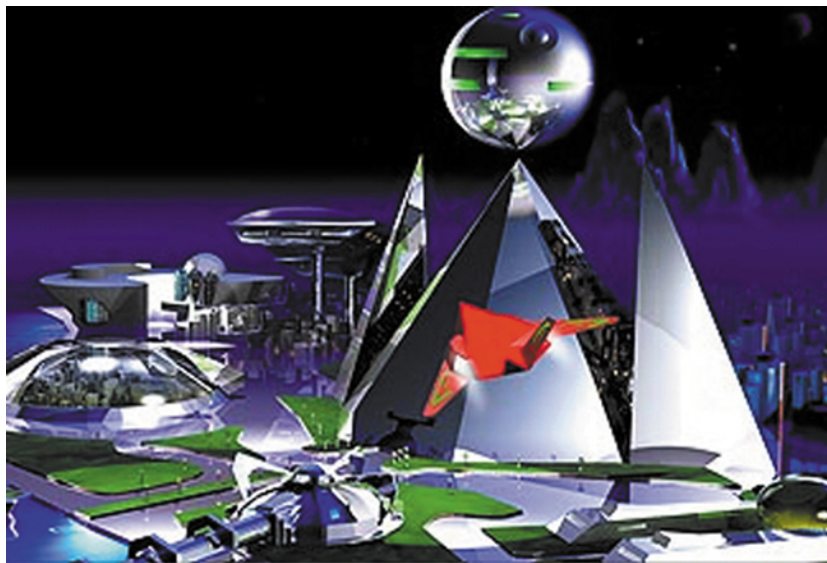


Fig. 2.4: Cybertown [cyb00b]

Avatars ab.

Neben 2D-Avataren sind im Cyberspace 3D Avatare zu finden (siehe Abbildung 2.5 von [ava00a]). Mit Hilfe dieser Avatare kann der Benutzer sich in Echtzeit durch virtuelle Welten bewegen und sich durch Gestik und Mimik ausdrücken. Ziel ist eine Erhöhung des Immersionsgrades durch einerseits möglichst realistisches Aussehen der Avatare sowie andererseits möglichst menschliches Verhalten. Dabei wird berücksichtigt, dass sich der Mensch nicht nur verbal, sondern auch non-verbal über Gestik, Mimik und Körpersprache ausdrückt.

Benutzerrepräsentierende Avatare werden hauptsächlich zum Zweck der Kommunikation eingesetzt, womit verschiedene Ziele erreicht werden. In virtuellen Universitäten zum Beispiel treffen sich Studenten zur Wissensbildung (vgl. [Con00]). Im einem virtuellen Spielcasino (vgl. [ava00b]) hingegen stehen für den Benutzer andere Aspekte im Vordergrund, zum Beispiel Unterhaltung. Von einigen Unternehmen werden diese virtuellen Personen auch zu Trainingszwecken eingesetzt, beispielsweise in [oSC00]. Der Benutzer betritt in Form eines Avatars virtuelle Trainingsumgebungen, in denen Reparaturen, Notfälle, schwierige Abläufe oder gefährliche Arbeitsbedingungen etc. simuliert werden. Schwierige Einsätze in lebensbedrohlichen Umgebungen können so geübt werden, ohne sich dabei tatsächlich in Gefahr zu begeben.

Ein besonders gelungenes Beispiel dieser Art ist das EU-geförderte Projekt ETOILE [DGS01] (Environment for Team, Organisational and Individual Learning in Emergencies), welches Notfallsituationen in Kernkraftwerken und U-Bahn Stationen nachstellt und für den Benutzer über einen persönlichen Repräsentanten greifbar und erfahrbar macht, siehe Abbildung 2.6.

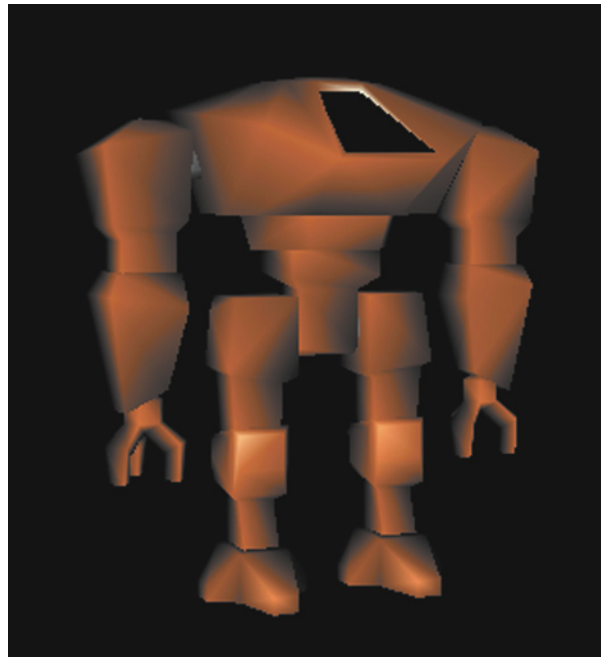


Fig. 2.5: 3D Avatar [ava00a]



Fig. 2.6: Sicht des Benutzers in virtueller Trainingsumgebung ETOILE (etoile.tecnatom.es)

Avatare dienen jedoch nicht ausschließlich als Manifestation von menschlichen Systembenutzern. Sie werden auch als Assistenten eingesetzt. Anwendungsbeispiele für Assistenten werden im folgenden Kapitel diskutiert.

2.1.2 Virtuelle Personen

Virtuelle Personen sind Avatare, die über einen eigenen 'menschlichen' Lebenslauf, Beruf etc. verfügen. Diese Personen treten inzwischen als Fernsehmoderatoren [ana00], Models [nod00], Sänger [e-c00] oder einfach als 'virtuelle Idole' auf. Die Modelagentur noDNA in Deutschland vermittelt ausschließlich virtuelle Models.

Bei der Realisierung menschlicher, virtueller Persönlichkeiten wird großer Wert auf lebens-echte Animation und realistisches Aussehen gelegt. Es muß sich bei dieser Art von Avataren jedoch nicht unbedingt um menschenähnliche Darstellung handeln. Als Beispiel sei hier Cleo (siehe Abb. 2.7) angeführt, die ihre eigene Sendung [cyb00a] im französischen Pay-per-View-Kanal Canal+ moderiert.



Fig. 2.7: Moderatorin Cleo von Canal+ [cyb00a]

2.1.3 Avatare als Assistenten

Diese Art von Avataren ist Gegenstand von Forschungen im Bereich Benutzeroberflächen bzw. Mensch-Maschine-Interaktion, die sowohl psychologische als auch technische Aspekte behandeln, vgl. [Wil97], [Tho96] und [WSS94].

Assistenzavatare sind in dieser Arbeit wie folgt definiert: Assistenzavatare sind eigenständige virtuelle Persönlichkeiten mit unterschiedlicher Intelligenz, die mit dem Benutzer kommunizieren und auf seine Anfragen reagieren. Sie werden nicht vom Benutzer gesteuert, sondern von einem bestehenden System. Ein Assistenzavatar hat eine oder mehrere bestimmte Aufgaben zu erfüllen. Diese können unter anderem im Bereich Datensuche und Aufbereitung, Hilfe bei wissensintensiven Anwendungen und generelle Informationspräsentation z.B. als Produkt- oder Unternehmenspräsentation liegen.

Als Teil der Benutzungsoberfläche werden Avatare durch weitere Aspekte definiert, auf die

in der Folge näher eingegangen wird:

Verhalten Avatare bieten die Möglichkeit, die Mensch-Maschine Kommunikation nicht nur durch geschriebenen oder gesprochenen Text durchzuführen, sondern zusätzliche Kommunikationskanäle einzubeziehen. Durch die Berücksichtigung der nonverbalen Kommunikation, die durch Mimik, Gestik und Körpersprache des Avatars erfolgt, wird der Dialog zwischen Mensch und Maschine *menschlicher*.⁶

Aussehen Avatare müssen nicht unbedingt anthropomorph⁷ gestaltet werden. Eine Auswahl zwischen tierischen und sächlichen Eigenschaften sowie zwischen verschiedenen Realitätsstufen ist anwendungsabhängig⁸ zu prüfen. Entscheidend ist, was der Benutzer dabei empfinden bzw. damit assoziieren soll, vgl. [WSS94].

Intelligenz Der dritte Aspekt behandelt die *Intelligenz*⁹ des avatarpräsentierten Systems. Folgende Beispiele sollen die Unterschiede verdeutlichen.

- Der Flaschengeist auf der Web-Seite des Mineralwasserherstellers Apollinaris, vgl. [apo00], kann einen sehr einfachen schriftlichen Dialog führen.
- Der Avatar Sylvie, vgl. [Per00], verarbeitet die vom Benutzer eingegebenen Texte oder Textdateien und gibt sie mittels Text-to-Speech aus. Sylvie hat eine explizite Aufgabe, nämlich vorlesen. Darüber hinaus kann auch sie nur sehr eingeschränkte Dialoge führen.
- Die 3D Assistenten des Projektes WebPersona, vgl. [ARM97], agieren auf Web-Seiten und können als Moderatoren auch für fremde Anwendungen fungieren. Ihre Intelligenz ist durch Anwendungsanbieter praktisch beliebig erweiterbar.

⁶ Dies führt zu den Grundsätzen der zwischenmenschlichen Kommunikation (Kommunikationsaxiome von Watzlawick, vgl. [Hei91]). Es besteht die Forderung, dass diese Axiome in die Interface-Planung einfließen sollten, um glaubwürdige Agenten entstehen zu lassen, vgl. [KK97]. Eines der Kommunikations-Axiome von Watzlawick beschreibt das unmögliche Unterfangen, explizit nicht zu kommunizieren. Übertragen auf avatarbasierte Interfaces bedeutet das, dass der Avatar dem Benutzer auch dann etwas mitteilt, wenn er gerade nicht spricht und eventuell auf Benutzereingaben wartet. Man spricht hierbei vom *single rest-pose problem*, vgl. [KK97]. Weitere Aspekte in der menschlichen Kommunikation sind das Verhältnis zwischen Sender und Empfänger sowie persönliche Erfahrungen.

⁷ Der Einsatz menschlicher Avatare, die sehr realistisch aussehen, ist nicht unproblematisch. Gesichter spielen in der menschlichen Entwicklung eine große Rolle. Mit menschlichen Gesichtern wird automatisch menschliches Verhalten assoziiert. (Das bedeutet, dass an einen lebensechten Avatar höhere Erwartungen bezüglich Intelligenz und Kommunikationsfähigkeit gerichtet werden als zum Beispiel an ein Cartoon-Gesicht, vgl. [WSS94].) Als zusätzliche Schwierigkeit verbindet jeder Mensch eigene Erfahrungen mit bestimmten Gesichtern.

⁸ Der Einsatz menschlicher Eigenschaften garantiert noch keine Akzeptanz. So akzeptiert z.B. der Benutzer eine anonyme, kühle, maschinenartige Fehlermeldungen leichter als solche, die ein menschähnlicher Avatar überbringt.

⁹ Dies führt über fest programmierte Animationssequenzen (wie zum Beispiel die Begrüßung durch einen Avatar) über gewisse Branching-Ansätze (je nach Benutzerverhalten wird eine bestimmte, immer noch fest programmierte Avatarverhaltenssequenz gezeigt) bis hin zu mittels künstlicher Intelligenz ausgewähltes, generierbares Avatarverhalten, vgl. Kapitel 3.2.

Die Vision für ein avatarbasiertes Assistenzsystem ist ein vom Aussehen her frei wählbarer Avatar, der auf dem Desktop quasi *lebt* und für die verschiedensten Aufgaben eingesetzt werden kann. Das generelle Ziel dieser Anwendung ist es, dass künftig die Hersteller von Anwendungen ihre Tools auf den 3D Assistenten ausrichten und dieser so als Interface für eine immer größere Zahl von Systemen zur Verfügung steht.

2.2 Assistenz-/Agentensysteme

In der Literatur werden die Begriffe *Assistent* und *Agent* kontrovers verwendet. In einigen Fällen werden die Begriffe synonym verwendet, in anderen werden Assistenten als Unterkategorie von Agenten betrachtet. Im folgenden wird der Begriff des Assistenten definiert:

Assistenten bzw. Assistenzsysteme sind Software-Tools, die dem Benutzer mittels spezialisierter Agenten, oftmals visualisiert mittels eines personalisierten Avatars, verschiedenste Aufgaben abnehmen können. Sie helfen unter anderem dabei, Informationen zu ordnen, zu sortieren und zu verarbeiten.

In dieser Arbeit wird ein Assistent als Unterkategorie von Agent interpretiert, d.h. als Agent mit der Bestimmung dem Benutzer durch Assistenz zu dienen. Allgemeine Agenten müssen nicht unbedingt direkten Kontakt zum Benutzer besitzen - sie können auch ausschließlich unter anderen Agenten arbeiten. Software-Agenten sind in der entsprechenden Literatur unterschiedlich definiert. Es lässt sich jedoch folgender Konsens für den Begriff Agent erkennen, welcher durch Wooldridge, vgl. [WJ95], notiert ist :

- Autonomie: Ein Agent ist in gewissen Grenzen zu selbstständigem Handeln befähigt.
- Proaktivität: Um gewisse Zielvorgaben zu erreichen versucht der Agent aktiv an den Zielen zu arbeiten.
- Reaktivität: Der Agent kennt seine Umgebung und kann in gewissen Grenzen darauf reagieren.
- Kommunikativität: Ein Agent kann mit anderen Agenten kommunizieren, um seine Ziele zu erreichen.

Im Hinblick auf ihre Eigenschaften sind verschiedene Kategorien von n zu unterscheiden:

- Intelligente Agenten: Diese Systeme basieren auf künstlicher Intelligenz, die zur Lösung von komplexen Aufgaben genutzt wird.
- Mobile Agenten: Diese Agenten sind nicht nur auf einer einzelnen Maschine tätig, sondern können ihren Ort der jeweiligen Aufgabenstellung anpassen und damit Ressourcen-orientiert ihre Aufgaben erledigen. Beruht ihre Intelligenz auf der Verteilung mehrerer einfacher Einheiten auf verschiedene Rechner, so spricht man auch von verteilten Agenten.

- Personalisierte Agenten: Agenten, welche aufgrund ihrer direkten Zuordnung zu einem spezifischen Benutzer diesen effizient in seiner Arbeit unterstützen. Personalisierte Agenten können i.A. durch die Verwendung von benutzerspezifischen Profilen die an sie gestellten Aufgaben mittels der vorhandenen Hintergrundinformationen (Interessen, Aufgaben und Möglichkeiten des Benutzers) in einem gewissen Rahmen selbstständig bzw. pro-aktiv (daher nicht direkt vom Benutzer beauftragt) erfüllen.

Doyle, vgl. [Doy99], kategorisiert den Einsatz kommunikativer Agenten in verschiedenen Anwendungsbereichen, abhängig vom Benutzerverhalten, nach folgenden Faktoren:

Entscheidungsfähigkeit des Benutzers Ein Benutzer, der eine Auswahl treffen muss und noch unsicher darüber ist, wird den Einsatz eines kommunikativen Agenten begrüßen. Dieser fungiert als Berater, der die für die Entscheidung erforderlichen Informationen liefert. Mögliches Einsatzgebiet: Lernsoftware.

Erwartungen des Benutzers Einsetzbar etwa im Unterhaltungsbereich, wo es für den Benutzer wichtig ist, dass ein Resultat zufriedenstellend ist, aber dessen Richtigkeit keine so große Rolle spielt.

Eingabemöglichkeiten Diese Entscheidung beim Interface-Design hängt vom Verständnis des Users für die Anwendung ab. Durch Dialog kann ein unsicherer Anwender Erklärungen bzw. Instruktionen erhalten. Ein erfahrener Benutzer wird mehr Wert auf eine reibungslose und rasche Abwicklung seiner Tasks legen und daher Kommandosteuerung bevorzugen.

Kommunikationsfähigkeit des Agenten Unterscheidung zwischen kommunikativen Agenten und solchen, die Sprache als Kommunikationsmittel verwenden. Von einem kommunikativen Agenten erwartet man soziale Kommunikation bzw. Konversation.

Eine Bewertung von Agentensystemen kann somit sowohl aus Sicht der Funktionalität der Agenten als auch aus Sicht der Möglichkeiten des Benutzers erfolgen.

2.2.1 Beispiele einfache Assistenzsysteme

Ziel der Anbieter von Assistenzsoftware ist es, möglichst viele Personen aus der Zielgruppe an ihre eigenen Produkte zu binden. Dazu ist es aufgrund des ansteigenden Konkurrenzdrucks wichtig, dass sich der Anbieter aus der Masse der anderen abhebt, vgl. [ARM97]. Dieses Ziel kann über die beiden Ausprägungen von Assistenzsystemen erreicht werden, nämlich einerseits der Funktionalität und andererseits der Benutzungsoberfläche.

Der Zusammenhang zwischen diesen beiden Aspekten kann speziell im Hinblick auf den Einsatz von Avataren beschrieben werden, vgl. [PKSW96]. Die Akzeptanz eines Assistenzsystems – also die Bereitschaft der Zielgruppe, es regelmäßig zu benutzen – ist stark von der Übereinstimmung zwischen angebotener Funktionalität und Oberfläche abhängig. Im folgenden wird eine Auswahl dieser einfachen Systeme beschrieben.

Norton HelpDesk Assistant Der Assistent (vgl. [sym00]) ermöglicht ein einfaches und effizientes Remote-Troubleshooting für Systemadministratoren. Im Zeitalter der Globalisierung, in dem weltweite Netzwerke aufgebaut werden, die Dezentralisierung von Unternehmen voranschreitet und Begriffe wie Teleworking und Telelearning an Bedeutung gewinnen, birgt die Systembetreuung ein hohes Aufwandspotential in sich. Der Norton HelpDesk Assistant soll die Administration innerhalb von Netzwerken einfacher und damit kostengünstiger machen.

Clean-Install Assistant Der Assistent, vgl. [mar00a], soll dem Mac OS-User die Systemverwaltung erleichtern. Der durchschnittliche Benutzer erhält dadurch eine wichtige Hilfestellung, die ihm den Erwerb und die intensive Beschäftigung mit Benutzerhandbüchern für Backup-Details etc. erspart. Immer komplexere Betriebssysteme machen eine solche Unterstützung bald unerlässlich.

Advisors Assistant Dieser Assistent, vgl. [cli00], wurde entwickelt, um Kundendaten zu verwalten. Er wurde speziell für Versicherungsunternehmen, Finanzberater etc. konzipiert. Das Tool soll eine effiziente Unterstützung im Umgang mit der explodierenden Menge an Kundendaten und damit eine nicht unerhebliche Kosteneinsparung bieten.

Desktop Assistant Hilfe für Kleinunternehmen in finanziellen Fragen bietet der Desktop Assistant, vgl. [ask00]. Er hilft bei der Interpretation von Finanzdaten, warnt bei kritischen finanziellen Verhältnissen etc. Der Desktop Assistant stellt einen automatischen und nur einmal zu bezahlenden Finanzberater dar.

Software Specification Assistant Dieser Assistent ist ein PC-basiertes Tool, vgl. [ccc00], das Informationen über Software-Anforderungsspezifikationen liefert. Diese Spezifikationen werden erstellt, bewertet und bei Bedarf aktualisiert.

RTI Kiosk Assistant Am Research Triangle Institute in North Carolina, USA, wurde ein avatarbasiertes Präsentationssystem mit wählbaren Charakteren entwickelt, das speziell auf Kioskpräsentationen ausgerichtet ist, siehe Abb. 2.8.

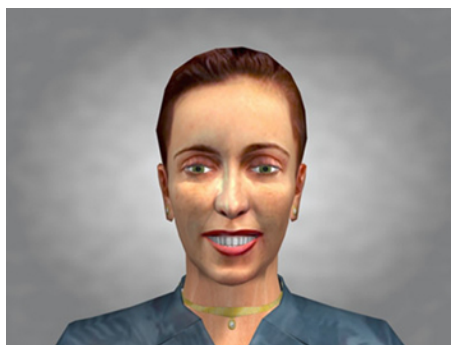


Fig. 2.8: Präsentationsavatar Roxanne von RTI

ZGDV Kiosk Assistant Am ZGDV (Zentrum für Graphische Datenverarbeitung) wurde ein Kiosk System, vgl. [Spi00], entwickelt, welches zwei Assistenten - einen virtuellen Avatar und einen physischen Roboter - zur Unterstützung des Benutzers bei der Auswahl von Themen realisiert. Hat der Benutzer Themen seines Interesses gefunden, so präsentieren die Kiosk-Assistent die Themen gemeinsam in einer multimedialen Weise, siehe Abbildung 2.9.



Fig. 2.9: Kiosk Assistenten, ZGDV

2.2.2 Komplexe Assistenzsysteme

Komplexe Assistenzsysteme repräsentieren i.A. eine Gruppe von Agenten bzw. einen speziellen, komplexen Agenten. Im Folgenden werden einige dieser Assistenzsysteme beschrieben.

Office 2000 Im Projekt Office 2000 des Fraunhofer-IGD, vgl. [BBL⁺01], ist ein Agenten-System, welches durch Assistenzfunktionen die Arbeit des typischen Büroanwenders entscheidend erleichtert. Im Zentrum der Applikation stehen zwei unterschiedliche Bereiche: So wird ein spezieller Arbeitstisch entwickelt, welcher die Anforderungen berücksichtigt, die eine veränderte Büroumgebung mit sich bringt. Zum anderen ist ein komplexes System von Assistenten realisiert, welches eine Delegation von Aufgaben durch den Benutzer ermöglicht, siehe Abbildung 2.10. Jeder Assistent wird dabei durch einen einzelnen Avatar dargestellt, um so das direkte Ansprechen eines Assistenten als auch eine Kontrolle über Arbeitsablauf und Arbeitsergebnisse über ein Agenten-spezifisches Interface zu erlauben.

WebPersona Im Rahmen des WebPersona Projekt des Deutschen Forschungszentrums für

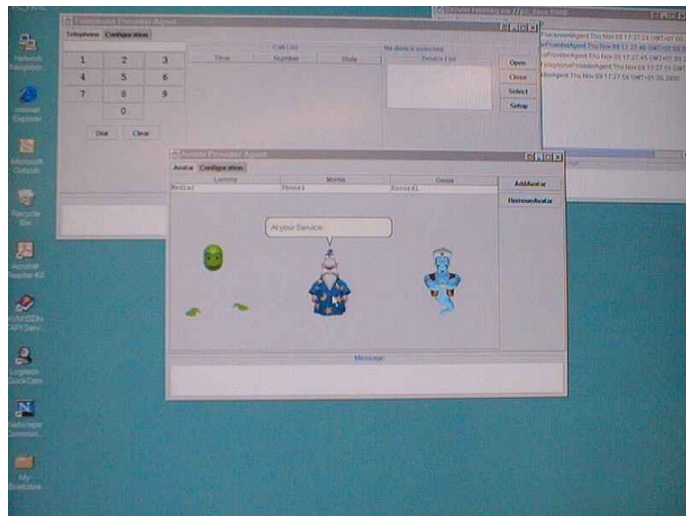


Fig. 2.10: Screenshot des Office 2000, Fraunhofer IGD

Künstliche Intelligenz (DFKI) werden Präsentationsavatare für das World Wide Web entwickelt, vgl. [ARM97]. Jeder Assistent zeichnet sich durch seine spezifischen Möglichkeiten der Gewinnung, Organisation und medialen Präsentation von Information aus. Der Assistent wird dabei als persönlicher Sekretär des Benutzers dargestellt.

2.3 Interaktive Narration und Digitales Video

Industrielle Anwendungen von interaktiver Narration und als Spezialfall Digitales Video finden sich äußerst spärlich.

Digitales Video ist in kommerziellen Produkten nicht automatisch mit interaktivem Video gleichzusetzen. In dieser Arbeit ist der Einsatz von interaktivem Video von besonderem Interesse. Interaktivität kann dabei verschiedentlich interpretiert werden, vgl [Rie99]:

Makrointeraktivität sind alle möglichen Aktivitäten, die das Video als Ganzes betreffen.

Unter diesen Punkt fallen zum Beispiel die Möglichkeit, ein Video beliebig zu wiederholen, unbedingte oder bedingte Verzweigungen zwischen einzelnen Videosequenzen, Festlegung der Reihenfolge von Videosequenzen.

Mikrointeraktivität sind alle Möglichkeiten, die Informationsdarbietung innerhalb einer einzelnen Videosequenz zu manipulieren, so zum Beispiel Änderungen der temporalen (Zeitlupe, Pause, Zeitraffer, Zeitsprung) oder spatialen (Zoom In, Zoom Out) Darstellungen der Videosequenz, Perspektiven- und Kamerawahl, Annotationen (hot spots).

Diese Interpretation ist weitgehend von der traditionellen Betrachtung linearer Videos abgeleitet und direkt inspiriert von der Art und Weise mit Videocassetten-Recordern oder

Videodisk-Playern umzugehen. Der in dieser Arbeit favorisierte Ansatz bezieht sich mit der Interaktion im wesentlichen auf die *Videoinhalte*. Diese Interaktion kann in die folgenden Kategorien¹⁰ unterschieden werden:

Synchrone Interaktion Die synchrone Interaktion bezieht sich auf die Inhalte (Inhaltsobjekte in ihrer graphischen und/oder akustischen Darstellung), welche zum Zeitpunkt der Interaktion im Video durch das Publikum erfahrbar sind.

Asynchrone Interaktion Die asynchrone Interaktion beinhaltet alle Interaktionen, welche sich nicht auf den aktuellen Präsentationszeitpunkt und die zum Präsentationszeitpunkt erfahrbaren Inhaltsobjekte beziehen.

Kommerzielle Anwendungen setzen Video im Internet bzw. Intranet in folgenden Bereichen ein:

- Präsentationen von Unternehmen, Vereinen, Organisationen etc.
- Produktpräsentationen, Online-Kataloge
- Kundenservice
- News
- Information
- Bildung
- Bewerbungen, Selbstdarstellung
- Unterhaltung, Musik
- Videohandel

Die Präsentation von Videos erfolgt dabei im Wesentlichen linear - eine Interaktion im direkt manipulativen Sinne erfolgt als sogenannte Makro-Interaktion. Im Folgenden ist eine Auswahl entsprechender kommerzieller Produktionen benannt:

Keller Products Bei diesem Beispiel handelt sich um Produktwerbung mittels Video. Die Firma Keller Products, vgl. [kel00] nützt die Möglichkeiten von Online-Video, um innovative Produkte in Form von 3D-Animationen zu beschreiben und Details darzustellen. Es handelt sich in diesem Fall um Büromöbel mit verschiedenen Features. Das Objekt wird zwar von allen Seiten und auch von innen abgebildet, es wird jedoch

¹⁰ Dieser Kurzdefinition folgt eine erweiterte, detaillierte Definition in Kapitel 4.3.2.

nur ansatzweise von der Möglichkeit Gebrauch gemacht, das Produkt während der tatsächlichen Verwendung darzustellen¹¹.

Plastic Plus Inc. Diese Online-Präsentation zählt ebenfalls zur Produktwerbung. Gezeigt wird ein neues Schneidsystem (Safetycutt Blade System, siehe [pla00]) für Rasenmäher. Die Aufnahmen zeigen das Produkt im Gebrauch, heben seine Vorzüge hervor und vermitteln innovative Aspekte. Unterstützt wird dies durch die Information, die der Sprecher transportiert. In diesem Fall wird das Medium Video linear eingesetzt.

Washington Post Die Washington Post, vgl. [was00], unterhält ein Videoarchiv. Die darin enthaltenen Videos dienen der Weiterbildung im Bereich Kochen. Grundlegende Handgriffe und Abläufe sollen mittels Videopräsentation leicht nachvollziehbar werden. Das Videobild wird durch gesprochene Erklärungen ergänzt.¹²

Distance Education and Technology Unter dem Titel *Distance Education and Technology*, vgl. [det00], bietet die University of British Columbia Videos zu verschiedenen Themengebieten¹³ an.

Auswahl von Bildungseinrichtungen Bei der Auswahl von Bildungseinrichtungen in Amerika hilft eine Videobibliothek, vgl. [sea00]. Schulen und Universitäten haben die Möglichkeit, ein Präsentationsvideo dort zu veröffentlichen und so für sich zu werben. Diese Videos liefern Informationen und bilden für diesen Zweck eine Ergänzung zu Broschüren etc.

ABC News Online Die Australian Broadcasting Corporation, vgl. [Aus00], zeigt Videos als Ergänzung der Online-Nachrichten. Durch die Präsentation von News-Videos im Internet sind Informationen – im Gegensatz zum Fernsehen – früher verfügbar und gegebenenfalls auch öfter abrufbar¹⁴.

Court TV Court TV, vgl. ([cou00]), bietet dem Internetbenutzer News-Videos an. Wie der Name des Fernsehsenders schon sagt, sind Neuigkeiten über aktuelle Gerichtsverhandlungen das Spezialgebiet. Court TV bietet alle aktuellen Videos in zeitlicher Reihenfolge an. Durch interaktives Video könnten die Beiträge nach Fällen geordnet strukturiert werden.

¹¹ Interaktivität in Video-Produktpräsentationen ist eher bei komplexen Produkten sinnvoll, zum Beispiel beim Verkauf von Autos. Durch das Video kann der Benutzer das Gesamtprodukt im Einsatz betrachten. Der Benutzer kann mittels Interaktion Detailinformationen abrufen, die bei der Kaufentscheidung eventuell eine Rolle spielen, jedoch zu speziell sind, um sie ohne Nachfrage des Benutzers zu zeigen. Angeboten wird so die gesamte Information über ein Produkt, der Konsument entscheidet aber selbst, ob er die angebotenen Daten tatsächlich benötigt und verlangt.

¹² Für die Zubereitung komplizierter Gerichte oder Menüs, ebenso wie für alle komplexen Abläufe, kann interaktives Video durchaus hilfreich sein, da man dadurch Informationen für alle Wissensstufen anbieten kann. Die Gesamtinformation wird in Teilabläufe aufgeteilt, die man je nach bereits vorhandenem Wissen ansehen oder überspringen kann.

¹³ Da Videos Informationen in Bild und Ton vermitteln, sind sie für die Wissensbildung generell gut einsetzbar, vor allem dann, wenn kein direkter Vortrag möglich ist.

¹⁴ Interaktivität in Videos ermöglicht den Zugriff auf Hintergrunddaten bzw. weiterführende Informationen über das Nachrichtenvideo. So kann beispielsweise über die Zeit zu einem Thema ein umfassendes Informationspaket bereitgestellt werden.

Interaktive Narration findet sich auch im Bereich Games und Entertainment. An dieser Stelle wird eine weitere Kategorisierung von Games vermieden und auf die verschiedenen Kategorisierungen in der Literatur, z.B. durch Crawford, vgl. [Cra82], hingewiesen. Die Narration von Geschichten in Spielen wird i.A. durch eine proprietäre, Applikationsspezifische Weise gelöst - generische, narrative Ansätze zum interaktiven Erzählen von Geschichten finden sich in der Spieleindustrie nicht - vor allem deshalb, weil die Relevanz von Narration in Games noch nicht durch die breite Masse der Entwicklungen in diesem Bereich erkannt wurde. Dieser Aspekt ist tatsächlich durch Laurel, vgl. [Lau93], in folgender Weise beschrieben: *Characters are lame in most boy games (...) they are so lame you can't even make up an interesting story about them.* Auf Szenen-Ebene findet in Games allerdings interaktive Narration statt. Dort werden i.A. Szenen-relevante Modellierungen als automatisierte narrative Elemente¹⁵ genutzt.

2.4 Zusammenfassung

In diesem Kapitel werden kommerzielle Anwendungen diskutiert, welche sich in Teilaspekten mit der Thematik der interaktiven Narration auf der Basis von Video beschäftigen. So werden sowohl Avatare, Assistenzsysteme als auch Digital Video Anwendungen diskutiert.

Die Einsichten dieses Kapitels zeigen, dass offenbar auf dem derzeitigen Markt für Videoanwendungen bzw. Anwendungen mit konversationaler Interaktion mittels Avataren kein Werkzeug existiert, welches speziell auf die Problematik kontextgetriebener, medienzentrierter, nichtlinearer Geschichtsnarration mit Video zugeschnitten ist. Es ist allerdings festzustellen, dass ein Bedarf an entsprechenden Systemen - sei es in Teilaspekten oder als Ganzes - besteht. An dieser Stelle wird deshalb die Forderung nach einem generischen System gestellt, das die interaktive Narration sowohl in der Modellbildung von Storys als auch in dem konversationalen Aspekt der Beeinflussung von Storys¹⁶ - und auch im Bezug auf allgemeine Assistenzsysteme als Seiteneffekt dieser Arbeit - unterstützt.

¹⁵ Die Thematik wird weiter vertieft in Kapitel 3.1.2.

¹⁶ Um diese Aspekte weiter zu vertiefen wird in Kapitel 3 eine Analyse zum Thema Nichtlineare Storytelling Systeme durchgeführt, welche mit einem Anforderungskatalog für das gewünschte System abschließt. Zur Vertiefung von Interaktionsmöglichkeiten des Publikums wird in Kapitel 4 die Interaktion auf kontinuierliche Medien analysiert. Die zu verwendenden Techniken und die Konzeption dieses Ansatzes werden in Kapitel 5 diskutiert, in den anschließenden Kapiteln Implementierungsvariationen vorgestellt und validiert.

3. NICHTLINEARE STORYTELLING SYSTEME

Erzählen ist gleichzusetzen mit der Notwendigkeit der Strukturierung von Informationen, um diese über einen Autor an ein Publikum weiterzugeben. Schon in einem sehr einfach strukturierten Fall der Kommunikationsforschung, der sogenannten *einseitigen Kommunikationskette*, werden Informationen von einem Sender zwangsläufig in einem gewissen Maße strukturiert und in einen Zusammenhang mit Bezugspunkten für den Empfänger gebracht, zum Empfänger gegeben. Der Empfänger hat in diesem einfachsten Fall keine Einflussmöglichkeit auf die Art der Darbietung der Information - der Informationsfluss ist unidirektional. Diese klassische, lineare Geschichte, die ausschließlich durch den Autor bestimmt ist, wird im Folgenden als Ausgangspunkt der Diskussion von nichtlinearer Narration gewählt. Im Anschluss werden - anhand der Einsichten, die bezüglich linearer Narration vorliegen - nonlineare Ansätze des Geschichtenerzählens untersucht. Anhand der nonlinearen Ansätze werden die aktuellen Konzepte der interaktiven Narration diskutiert und im Anschluss ein Anforderungskatalog¹ für die nichtlineare Narration von Video aufgestellt.

3.1 Lineares Storytelling

3.1.1 Grundzüge einer Geschichte

Jede Geschichte, unabhängig von Ihrem Inhalt, besitzt gewisse Eigenarten², die sie mit jeder anderen Geschichte gemeinsam hat, vgl. [LBSA91]:

Eine Geschichte bietet einen Kontext, innerhalb dessen gewisse Handlungen oder Aktionen durch Charaktere ausgeführt werden. Durch die Erzählung der Geschichte werden Raum und Zeit manipuliert, um beim Empfänger der Geschichte gewisse kognitive Prozesse in Gang zu setzen.

Die Art und Weise, wie eine Geschichte dann letztlich erzählt wird, macht aus einer Ansammlung von Informationen bezüglich Kontext, Charakteren und deren Handlung etwas Einzigartiges³.

Das Erzählen selbst in der Begrifflichkeit der *Narration* wird nach Bordwell, vgl. [Bor85],

¹ Dieser Anforderungskatalog dient als Basis zur Erstellung des in dieser Arbeit vorgestellten Konzeptes zur interaktiven Narration von kontinuierlichen Medien, siehe Kapitel 5.

² Diese Eigenarten werden als nicht medial angesehen - es ist nicht relevant, ob die Geschichte mittels Text, Bild oder Bewegtbild (Film, Video) ausgedrückt wird.

³ Ein Beispiel hierfür sind Tatsachenberichte durch Augenzeugen - jeder Augenzeuge gibt eine eigene Geschichte zum besten, obwohl jede der Geschichten auf den gleichen Tatsachen basiert.

in zwei Ausprägungen beschrieben:

Diegetische Narration Das tatsächliche Erzählen einer Handlung durch einen Erzähler, der als separiert von der Geschichte, d.h. nicht als in der Geschichte handelnde Figur, zu erkennen ist.

Mimetische Narration Das Zeigen eines Geschehens in einer dramatischen Handlung⁴. In diesem Fall spricht der Erzähler durch die Figuren der Geschichte.

Film, Fernsehen und Video sowie Medienpräsentationen im Allgemeinen verbinden die beiden Formen der Narration.

Es gibt einige Theorien des Geschichtenerzählens, die untereinander sehr ähnlich sind. Die vorliegende Arbeit orientiert sich an Mallon und Webb, vgl. [LMW97]:

*Die Erzählung ist eine Einheit, die Struktur, Inhalt, Kontext und Entwicklung integriert. Diese Einheit besitzt eine Reihe von Eigenschaften, wie ein globales und lokales Ursache-Wirkung-Netzwerk, eine Datenbasis mit komplexen, wechselseitigen Zusammenhängen, eine konstruierte Realität sowie das Fehlen einer allgemeingültigen Struktur*⁵.

Die meisten Geschichten halten sich an einen sehr einfachen Sonderfall des Erzählschemas: Die Heldengeschichte. Diese basiert auf einer mythischen Erzählstruktur, vgl. [Cam99]:

Der Heros verlässt die Welt des gemeinen Tages und sucht einen Bereich übernatürlicher Wunder auf, besteht dort fabelartige Mächte und erringt einen entscheidenden Sieg, dann kehrt er mit der Kraft, seine Mitmenschen mit Segnungen zu versehen, von seiner geheimniserfüllten Fahrt zurück.

Das Erzählschema der Heldengeschichte wird als Dramaturgiemodell auch durch Field, vgl. [Fie87], beschrieben, siehe Bild 3.1.



Fig. 3.1: Dramaturgie in 3 Akten nach Syd Field

Das Software Werkzeug *Dramatica*, vgl. [Dra], greift die oben geschilderten Theorien auf und entwickelt daraus eine Theorie des Geschichtenerzählens, welche charakteristische Elemente für die gängigen Art der linearen Geschichten beschreibt. Mit *Dramatica* wird die Geschichte geteilt in einen Geschichts-formenden und einen Geschichts-erzählenden Teil.

- Die Geschichtsformung beschäftigt sich mit den zu erzählenden Themen, mit der

⁴ Die Handlung wird im Aristotelischen Sinne als Nachahmung und Immitation, d.h. als Mimesis verstanden.

⁵ Die fehlende allgemeingültige Struktur von Erzählungen impliziert nach Meinung von Mallon und Webb den Ausschluß einer allgemeingültigen Meta-Geschichten-Maschine.

Psychologie der Charaktere usw.. Die entsprechenden Eigenschaften werden mit einer Menge von Fragen abgearbeitet und in Dramatica gespeichert.

- Beim Geschichtenerzählen werden die entsprechenden Charaktere quasi instantiiert, d.h. mit zusätzlichen Informationen spezialisiert. Dramatica achtet dabei peinlich genau darauf, dass die Entwicklung der Charaktere in der Geschichte tatsächlich deren definierter Form entspricht.

Die zentralen Konzepte der Dramatica Theorie orientieren sich am Begriff des Story Mind⁶, welches das Problem darstellt, das gelöst werden soll. Diese Lösung sollte möglichst umfassend, d.h. von allen verschiedenen Standpunkten und Sichtweisen her, beschrieben werden. The Four Throughlines, d.h. die 4 Hauptblickrichtungen auf ein Problem, sind die Objective Story Throughline, die Main Character Throughline, die Obstacle Character Throughline und die Subjective Story Throughline.

- The Story Mind: Ein holistisches Konzept, welches alle für eine Geschichte relevanten Daten beinhaltet, daher den gesamten Kontext, alle Charaktere, Handlungen, Thematik und das Genre der Geschichte.
- The Four Throughlines: Jede Geschichte wird aus einer bestimmten Perspektive vom Zuhörer wahrgenommen. Diese Perspektiven werden durch die Akteure einer Geschichte eingenommen. Ein Akteur kann auch eine Problematik sein, welche durch die Geschichte ausgedrückt wird. Diese Problematik wird als treibende Kraft der Geschichte und damit als Akteur gesehen. Folgende vier Perspektiven werden unterschieden:
 - The Objective Story Throughline: Aus der Perspektive des allwissenden Erzählers die Gesamtheit aller Informationen einer Geschichte, wie sie sich einem objektiven Betrachter offenbart. Dazu gehört in besonderem Maße die subjektive Sicht einzelner Akteure.
 - The Main Character Throughline: Der Main Character ist die Hauptperson der Geschichte und somit handelt es sich um die Sichtweise der Geschichte aus der Perspektive des Haupt-Akteurs oder des sogenannten Helden der Geschichte.
 - The Obstacle Character Throughline. In jeder Geschichte gibt es zumindest einen Konflikt, welcher im Verlauf der Geschichte überwunden werden muss. Die Perspektive dieser Problematik ist quasie die Sichtweise des Anti-Helden der Geschichte.
 - The Subjective Story Throughline: Die subjektive Sicht der Geschichte mit dem Focus auf den Kampf zwischen Held und Anti-Held.
- The Grand Argument Story: Die gut argumentierte Geschichte bietet eine logische Handlung mit einem Focus auf die vier Story Throughlines, welche es dem Zuhörer der Geschichte ermöglichen, die Geschichte kognitiv zu verarbeiten und komplett zu verstehen.

⁶ Die klassische Literaturtheorie kennt hierfür den Begriff der Fabula.

Die verschiedenen Bausteine der Geschichte werden benutzt, um ein Konfliktschema zu beschreiben. In diesem Schema setzt sich ein Protagonist mit einem Antagonisten auseinander. Deren Konflikt wird in einer Ansammlung von Handlungen und Ereignissen ausgetragen, welche zu einer dramatischen Auflösung führen. Durch die Kombination von physischer und emotionaler Handlung wird ein Interesse beim Zuschauer aufgebaut, welches dem Zuschauer eine Immersion in die Geschichte ermöglicht.

Die Grand Argument Story - d.h. eine Geschichte, die durch Beachtung der vier oben genannten Kriterien Ihre Story *gut argumentiert*, also die wichtigsten Aspekte der Story darstellt, ist eine vollständige Geschichte nach dem Konzept von Dramatica.

Mit den oben beschriebenen Bausteinen kann jede erdenkliche Geschichte beschrieben werden. Tatsächlich existieren sogar feste Regeln für den dramaturgischen Aufbau von Filmen, welche solche Geschichten erzählen. Diese Regeln orientieren sich an dem gebräuchlichsten Hilfsmittel, mit dem ein Film beschrieben wird: Dem Drehbuch.

Das Drehbuch ist die Gesamtzahl aller Szenen eines Filmes. Jede Szene ist dabei erschöpfend beschrieben, d.h. Ort, Zeit, Charaktere, Handlung. Mehrere Szenen können zu einer Sequenz zusammengefaßt werden. Jede Sequenz entspricht dabei gewissermaßen einer Idee, die filmisch umgesetzt wird. Die Sequenzen wiederum sind geordnet in Anfang, Mitte und Schluss, d.h. 3 Akten, was dem Schema von Field, vgl. [Fie87], entspricht.

Die Dramaturgie des Drehbuchs wird von verschiedenen Autoren als teilweise automatisierbar erachtet. Um dem Zuschauer einen gewissen *Suspense*⁷ zu ermöglichen und ihn an die Geschichte zu fesseln, d.h. eine Immersion zu erzeugen, wird u.a. von Field vorgeschlagen, dass in regelmäßigen Abständen von 10 bis 15 Minuten etwas *relevantes*⁸ im Film passieren sollte. Andere Experten, wie z.B. Vale, vgl. [Ell95], gehen von einem dramaturgischen Aufbau aus, der sich unabhängig von jeder Geschichte genau bestimmen lässt.

Der Formalist Propp [Pro58] geht über die allgemeinen Designvorschläge für Drehbücher etc. hinaus und bewertet eine Geschichte vollkommen unabhängig von ihrem tatsächlichen Inhalt nach der morphologischen Funktion von Inhaltsteilen der Geschichte⁹, wie u.a. durch Berger, vgl. [Ber00], erläutert.

Propp definiert die (konventionelle Begrifflichkeit der) Handlung als eine Reihe von Motiven und somit als sekundären kreativen Prozess. Der primäre Prozess des Autors besteht in der Nutzung von Funktionen, die jedes Motiv als eine Erzähleinheit (oder einen Narratem¹⁰, bzw. Morphem¹¹) abbilden. Die originale Definition von Funktionen nach Propp ist im folgenden zitiert¹²:

⁷ Als Suspense wird die Spannung bezeichnet, welche der Zuschauer beim erleben der Erzählung erfährt. Spannung wird vor allem durch eine Lücke zwischen der Erwartung des Zuschauers und der Auflösung des Autors erzeugt.

⁸ Als *relevant* wird hierbei etwas bezeichnet, was die Erzählung in eine gewisse Richtung vorantreibt.

⁹ Propp erstellt sein Geschichtenmodell anhand einer Analyse von russischen Volksmärchen - die Analyse kann jedoch auf alle Formen von Geschichten übertragen werden.

¹⁰ vgl. [Ber00].

¹¹ vgl. [Mur98].

¹² Vgl. [Pro58], S. 21.

Function must be taken as an act of a character, defined from the point of view of its significance for the course of the action.

Dies bedeutet im Wesentlichen, dass ein und dieselbe Handlung unterschiedliche Funktionen in einer Geschichte einnehmen kann - je nach Kontext und Zeitpunkt, daher der Anordnung innerhalb der Gesamtgeschichte. Eine limitierte Anzahl von Funktionen dienen als Basiseinheiten einer Geschichte, unabhängig davon welcher konkrete Charakter oder welche konkrete Aktion die Funktion letztlich ausführt. Charaktere werden als dramatische Rollen (sogenannte *dramatis personae*) aufgefasst, die eine gewisse allgemeine Gültigkeit in ihrer Rolle besitzen:

- Der Held¹³
- Der Feind
- Der Gönner/Wohltäter¹⁴
- Der magische Helfer¹⁵
- Der falsche Held¹⁶
- Die Prinzessin¹⁷
- Das Opfer¹⁸

Die Zahl der Funktionen ist beschränkt, ihre Abfolge kann durch ein Regelwerk spezifiziert werden, da die Abfolge im Wesentlichen immer gleich ist. Vorgezogene, sich wiederholende, ausfallende, variante, sequentielle, überlappende oder disjunkte Funktionen werden in dem Regelwerk explizit berücksichtigt. Die Gesamtheit der morphologischen Funktionen von Propp ist im Anhang A.1 dieser Arbeit aufgelistet.

Propp schildert zwei generelle, disjunkte Kategorien von Märchen:

- Sequenzen, in denen der Held den Feind in einem Kampf besiegt, siehe Abbildung 3.2.
- Sequenzen, in denen der Held eine schwierige Aufgabe löst, siehe Abbildung 3.3.

Ein Märchen kann aus einer oder mehreren Sequenzen bestehen, wobei jedoch die Kampf-Sequenz immer vor der Test-Sequenz erzählt wird. Jede Funktion kann in unterschiedlichsten Varianten vorkommen, so werden von Propp z.B. 19 Fälle eines Feindesanschlages (A) differenziert.

¹³ Der Held wird auch als Suchender oder Opfer bezeichnet.

¹⁴ Vom Gönner bekommt der Held einen magischen Helfer.

¹⁵ Der magische Helfer hilft dem Helden sein Problem zu lösen.

¹⁶ Der falsche Held beansprucht die Heldentaten des Helden für sich

¹⁷ Die Prinzessin steht stellvertretend für *das große Glück*.

¹⁸ Das Opfer ist eine Person (falls nicht der Held), welche durch den Feind angegriffen wird.

Fig. 3.2: Story Sequenz *Kampf mit dem Feind* nach ProppFig. 3.3: Story Sequenz *Lösung einer schwierigen Aufgabe* nach Propp

Häufig werden zu Beginn eines Märchens idyllische Verhältnisse geschildert, welche im Kontrast zu späteren Unglücken stehen und diese Unglücke zumindest erahnen lassen. So wird zur Abwesenheit der Eltern ein Verbot erzählt, das die Eltern vor ihrer Abreise ausgesprochen haben. Die Geschichte nimmt ihren Lauf, während dessen sich verschiedene Abhängigkeiten ergeben können:

- A → K: Ein verübtes Unrecht wird immer wieder gut gemacht.
- H → I: Ein Kampf gegen einen Feind führt notwendig zum Sieg.
- M → N: Eine schwierige Aufgabe wird immer bewältigt.
- Pr → Rs: Die Verfolgung des Helden führt stets zu seiner Rettung.
- L → Ex: Ein falscher Held wird immer enttarnt.

Durch diese Definition der morphologischen Funktionen im Kontext der gesamten Geschichte löst Propp die Aufgabe, eine Geschichte in diskrete Module zu zerlegen trotz der starken

Verknüpfung der einzelnen Geschichts-Elemente untereinander. Er folgt somit den gestalterischen Regeln von Mallon und Webb¹⁹, vgl. [LMW97]. Offenbar definiert Propp die Summe der Ereignisse als mehr denn deren einzelner 'Wert' für die Geschichte.

3.1.2 Spannung mittels Filmtechniken

Bei filmisch umgesetzten Geschichten werden tatsächlich mehrere Medien miteinander kombiniert, um Spannung zu erzeugen. Als erster Filmmacher hat dies Sergei Eisenstein erkannt, vgl. [Eis67], der eine eigene Notationstechnik für die Dramaturgie eines Filmes erstellte, siehe Bild 3.4. So definierte Eisenstein für die Filmmontage gewisse grundlegende Einheiten, die *Attraktionen*²⁰. Die einzelne Attraktion besitzt eine von vier charakteristischen Informationswirkungen²¹. Die Kombination der Attraktionen auf einem Zeitstrahl ergibt die Dramaturgie eines Filmes.

Für das Bewegtbild bestimmt die Kameraführung und der Film-Schnitt wesentlich, welche Emotionen beim Zuschauer erzeugt werden. Sie ist ein Ausdrucksmittel, das erst mit der Einführung von Film entdeckt wurde. Um diese grundlegenden Filmtechniken festzulegen, existiert beim Film ein spezielles Werkzeug, das Storyboard. Es zeigt für jede Szene die Anordnung der beteiligten Charaktere und des Kontextes aus dem Blickwinkel der Kamera. (In Kapitel 3.2.3 wird dieses Thema zur Visualisierung von Avataren in Szenen erneut aufgegriffen.)

Diese Narration findet sich in Grundzügen auch im Bereich der Spiele-Industrie. Dort werden i.A. Szenen-relevante Modellierungen als automatisierte narrative Elemente genutzt. Bates, vgl. [BS89], gibt für die grundlegende Filmtechnik verschiedene Einstellungen²² vor. Für die Kameraeinstellung²³ alleine ergibt sich ebenfalls eine emotionale Wirkung, welche u.a. von Spierling und Gaida, vgl. [SG00], beschrieben wird.

¹⁹ Notiert zu Beginn des Kapitels 3.2.

²⁰ Eine Attraktion ist entweder ein Bild, ein Bewegtbild, ein Text oder ein Ton.

²¹ Die Informationswirkung wird von Eisenstein notiert als definierend, beschreibend, erörternd oder assoziierend.

²² Die Einstellungen sind in Anhang A.3 mitsamt ihrer narrativen Wirkung aufgelistet.

²³ Diese klassischen Kameraperspektiven sind in Anhang A.3 beschrieben.

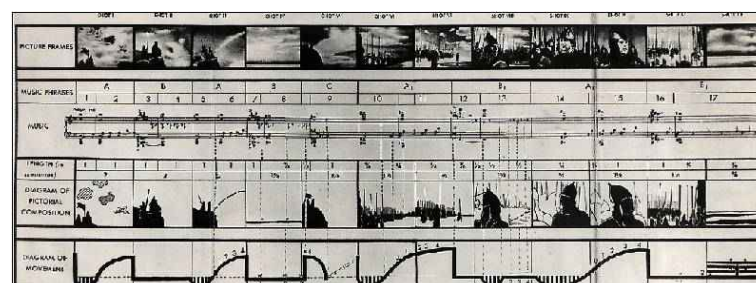


Fig. 3.4: Filmmontage nach Eisenstein [Eis67]

3.1.3 Zusammenfassung

Die Theorie der linearen Geschichte, speziell für Film und Video, kann nicht ohne eine gewisse Revision auf nichtlineare Geschichten übertragen werden - denn durch Interaktion kann der Zuhörer die Struktur, d.h. die Throughline der Geschichte, verändern - und damit die gesamte Dramaturgie sowie die dadurch beim Zuschauer erzeugte Immersion und Emotion. Damit übernimmt der Zuhörer einen gewissen Teil der Aufgaben des Autors. Je nach Komplexität des verfolgten, nichtlinearen Ansatzes kann der Zuschauer von der Auswahl verschiedener Storythroughlines bis hin zur Beeinflussung von Charakteren und damit der Schaffung benutzerspezifischer Storythroughlines die Geschichte beeinflussen.

3.2 Nichtlineares Storytelling

Das Erzählen von nichtlinearen Geschichten ist mit traditionellen Methoden natürlich vor Allem dann möglich, wenn der Autor und Erzähler der Geschichte explizit anwesend ist. Mit digitalen Medien und mit Hilfe von Computern stehen dem Autor diverse Möglichkeiten zur Verfügung, um eine Geschichte nichtlinear, d.h. bestimmt durch die Interaktion eines Publikums, zu erzählen, ohne dass sein direktes Eingreifen notwendig ist. Die dadurch entstehende Beeinflussung der Narration wird durch die Nutzung von digitalen Medien sowohl für den Autor als auch das Publikum ermöglicht.

Die nichtlineare Geschichte ist in dieser Arbeit somit wie folgt definiert: Eine nichtlineare Geschichte besteht aus einem Storymind und Throughlines, die an einer oder mehreren Stellen nicht fest vorgegeben und beeinflussbar sind. Die Begriffe Storymind und Throughlines entsprechen den Begrifflichkeiten der Dramatica Theorie.

Neben der Organisation der Geschichte und ihrer Elemente ist der gestalterische Aspekt von Geschichten von besonderem Interesse. So existieren zur narrativen Nutzung von digitalen Medien die folgenden Richtlinien, vgl. Linane-Mallon und Web [LMW97]:

- Strukturierung und Organisation von Inhalten: Zusammenhang und Integration der präsentierten Daten werden als Menge, Katalog, Episoden, ungerichtete oder gerichtete Kette, einfache Narration oder Narration²⁴ angegeben.
- Interface und Navigation: An dieser Stelle wird die Unsichtbarkeit²⁵ (Hiddenness) des Benutzerinterfaces gefordert.
- Konstruktion (Constructedness): Die Ausdehnung und der Charakter²⁶ einer Präsen-

²⁴ Die Narration entspricht einem idealen Zustand der Datenorganisation, in welchem Dateneinheiten auf einem Mikrolevel verbunden werden, was globale Folgen und Effekte für den gesamten narrativen Datenverbund besitzt.

²⁵ Dies hat einen direkten Bezug zur Immersionsfähigkeit des Benutzers: Je weniger auffällig das Interface, desto eher kann der Benutzer eine Immersion in die Geschichte aufbauen. Benutzertests zeigen jedoch, dass dies als *Hygienefaktor* eines Systems zu sehen ist: Ein unsichtbares Interface zwingt keine Immersion des Benutzers, jedoch behindert ein deutlich sichtbares Interface die Immersion des Benutzers.

²⁶ Ganz im Sinne einer gut konstruierten bzw. argumentierten Geschichte, siehe Dramatica in Kapitel 3.1.1.

tation.

- Räumlichkeit: Mit dieser Bezeichnung ist die räumliche²⁷ Ausprägung einer Geschichte gemeint.

Eine Ebene abstrakter ist die Kategorisierung von Barret, vgl. [Bar00], die Konsistenz, Plastizität, Persistenz, Skalierbarkeit, Komplexität, organische Details und Charakter Interaktion beinhalten. Nach Murray, vgl. [Mur98], sollte eine mit digitalen Medien erzählte Geschichte prozedural (nichtlinear), partizipatorisch (teilhabend), räumlich²⁸ und enzyklopädisch (vollständig) sein.

Faktisch ergeben sich die oben genannten Anforderungen direkt aus dem Maß der Interaktion, die dem Publikum geboten wird - je freier die Interaktionsmöglichkeiten des Publikums, desto umfassender muss die Geschichte gestaltet sein. Die Art der Partizipation bestimmt dabei wesentlich, wie die digitale Umgebung gestaltet wird. Folgende *Partizipationsmöglichkeiten* können unterschieden werden:

- Manipulation der Sichtweise auf die Geschichte in Zeit und Raum²⁹.
- Beeinflussung der Charaktere³⁰ einer Geschichte.
- Übernahme eines Charakters der Geschichte und damit Aktionsfähigkeit³¹ innerhalb der Geschichte.

Rein navigatorische Einflussnahme auf die Geschichte, daher das Ändern der Reihenfolge der Präsentation einer Geschichte, wird nicht als narrativer Einfluss auf die Geschichte gesehen.

Mit dem Wissen um die Ästhetik und den Benutzereinfluss auf Narration wird im Folgenden die Grundlage von interaktiver Narration notiert und interaktive Narrationsmodelle diskutiert.

3.2.1 Grundlagen

Eine Geschichte kann als Orchestrierung verschiedener Aktionen, ausgeführt von verschiedenen Akteuren innerhalb eines zeitlichen und räumlichen Kontextes gesehen werden. Lässt

²⁷ Diese Ausprägung ist jedoch auf die zeitliche Ausdehnung ebenso zu erweitern, insbesondere die Ausdehnung von Erzählzeit und erzählter Zeit, vgl. [Sut00].

²⁸ An dieser Stelle ist sicherlich noch zu fordern, dass eine Geschichte, vermittelt über eine digitale Umgebung, mit zeitlichen Aspekten versehen, d.h. temporal ist.

²⁹ Dies entspricht einer Sicht von Außen auf die Geschichte mit der Möglichkeit, den Standpunkt zu wechseln. Es ist somit eine Änderung der subjektiven Troughline einer Geschichte.

³⁰ Dies ist die Sichtweise von Innen, mit der Möglichkeit, die Geschichte direkt zu beeinflussen, z.B. durch Ändern der Eigenschaften eines Charakters, der Umgebung (Raum), der Zeit. Erzählt wird die Geschichte jedoch durch das digitale Medium. Dies ist eine Änderung der objektiven Troughline einer Geschichte.

³¹ Aktionsfähigkeit bedeutet in diesem Zusammenhang, dass das Mitspielen in der Geschichte ermöglicht wird. Der Benutzer wird selbst zum Mitautor der Geschichte. Dies entspricht der Änderungsmöglichkeit einer Main Character Troughline oder einer Obstacle Character Troughline.

der Autor einer Geschichte zu, dass durch eine Interaktion des Zuhörers mit dem Geschichtenerzähler diese Orchestrierung gestört wird, so muss es einen guten Grund dafür geben: Der Autor erwartet, dass der Zuschauer gesteigerte Aufmerksamkeit gegenüber dem Ziel der Geschichte entwickelt. Es ist sogar eine neue Art von Geschichten denkbar - Geschichten³², deren Ziel erst durch Interaktion des Zuschauers erreicht werden kann. Die Beeinflussung einer Geschichte durch den Zuschauer muss jedoch gezielt eingesetzt werden, um das Ziel des Autors tatsächlich erreichen zu können, denn der Autor steht bei Interaktivem Storytelling vor einem besonderen Dilemma.

Dieses Dilemma ist vor allem seit der Verbreitung von Hypertext und den damit verbundenen Möglichkeiten der Hypertext-Storys bekannt geworden: Wann weiß das Publikum, dass eine Geschichte tatsächlich *zu Ende*³³ ist bzw. dass es auf dem richtigen Weg durch die Geschichte ist?

Die Möglichkeiten der Interaktion stellen den Autor aus dem oben notierten Grund vor besondere Schwierigkeiten: Er möchte dem Publikum seiner Geschichte ein bestimmtes Ziel in Reichweite bringen. Dieses Ziel wird jedoch, auf Grund der mannigfaltigen Einflussmöglichkeiten durch das Publikum, eventuell nicht erreicht. Der Autor muss dem zu Folge Vorkehrungen treffen, um dem Publikum das Erreichen des Zieles zu ermöglichen ohne das Publikum in seiner Interaktionsfähigkeit zu sehr einzuschränken und damit die Immersion zu stören.

Dieses Dilemma existiert für jede Form des nichtlinearen Geschichtenerzählens. Um das Dilemma zu Überwinden, muss der Autor einer Geschichte sich über die folgenden drei grundsätzlichen, sich wechselseitig beeinflussenden Einflussgrößen, vgl. [Mat97] auf eine nichtlineare Geschichte im Klaren sein, siehe Abbildung 3.5:

Lokale oder Globale Kontrolle Eine Design-Entscheidung innerhalb von Narrationsansätzen ist die Ausprägung der Kontrolle von Narration. Diese kann lokal³⁴ in den Primitiven geschehen, wie zum Beispiel Spielfiguren (lokale Agenten oder Avatare). Branching oder Scripting-Ansätze bieten eine globalere, jedoch an der Vergangenheit der Erzählung ausgerichtete, Kontrolle³⁵. Eine globale Kontrolle ergibt sich, wenn aus der historischen Plot-Struktur alle zukünftig erreichbaren Plot-Strukturen in Betracht gezogen werden und aus diesen das Optimum ausgewählt wird.

³² Ein einfaches Beispiel dafür sind Lernspiele oder Gruppenunterricht. Erst durch die Interaktion der Lernenden mit dem Lehrer, dem Lernstoff und den anderen Lernenden ist das Lernziel zu erreichen und damit die Geschichte zu erleben.

³³ Ein sehr bekanntes Beispiel dieser Problematik wird anhand der Hypertext-Story *Afternoon* durch Walker, vgl. [Wal99], besprochen. Walker erkennt, dass die Suche nach dem Ziel bzw. nach dem Ende der Geschichte durch das Publikum ein nicht triviales Unterfangen ist. Letzten Endes weiß das Publikum erst, dass es das Ende der Geschichte erreicht hat, wenn es alle Hyperlink-Möglichkeiten ausprobiert hat.

³⁴ Mit der lokalen Kontrolle werden ausschließlich Aktionen von Moment zu Moment strukturiert. Dies ist für die Entwicklung einer Geschichte durch einen Autor unvorteilhaft, die Geschichte wird von einzelnen Interaktionen eher zufällig vorangetrieben.

³⁵ Diese Systeme beziehen Information aus der Erzählweise der bisherigen Narration (Historie der Narration), um daraus die Fortsetzung der Geschichte zu berechnen. Weiterhin kann die Narration an Bedingungen geknüpft sein, d.h. der nächste Plot der Geschichte wird erst angesteuert, wenn gewisse Eigenschaften erfüllt und Hindernisse überwunden sind.

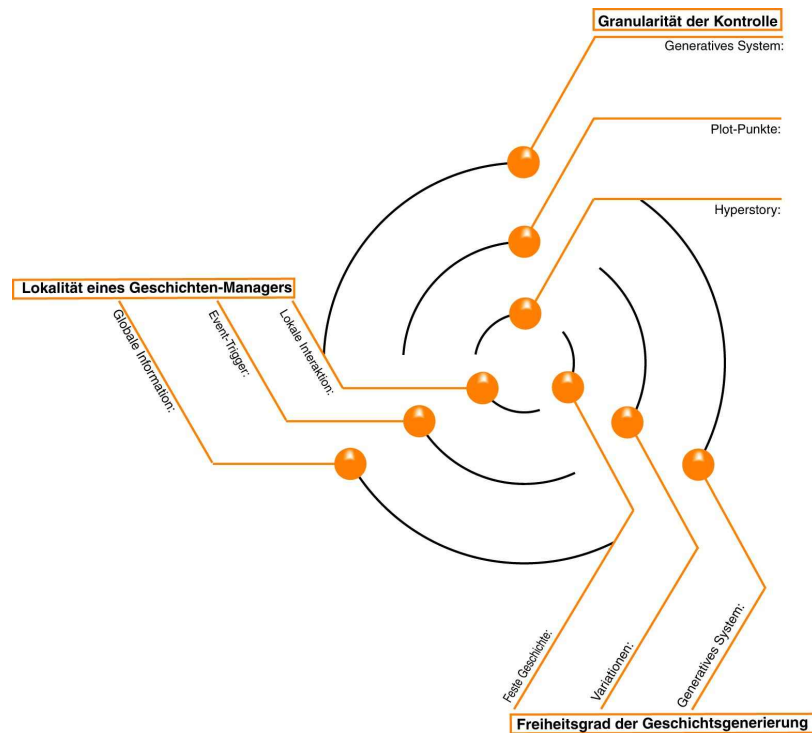


Fig. 3.5: Eigenschaften des narrativen Designs einer Geschichte

Granularität der Kontrolle Die Granularität der Kontrolle beginnt mit der sehr feinen Kontrolle aller Aktionen des Publikums und endet mit der groben Kontrolle der Geschichte ohne direkte Kontrolle der einzelnen Publikumsaktionen.

Eine sehr feine Kontrolle ist zum Beispiel in Hypermedia³⁶ Systemen gegeben. Dieser Ansatz ist vor allem dann vorteilhaft, wenn keine Ereignisse generiert werden können. Dies ist der Fall bei selektiver³⁷ Narration. Diese Narration wird *Lineare Geschichte mit interaktiven Elementen*³⁸ genannt, vgl. [LBSA91].

Eine mittelfeine Kontrolle ist in Systemen gegeben, welche die Szenen der Geschich-

³⁶ Hypermedia, siehe Definition Kapitel 4. Jede mögliche Publikumsinteraktion ist im Voraus vom Autor festgelegt. Die Verzweigungen im System geben präzise an, wann und wo innerhalb der Geschichte handlungsrelevante Ereignisse stattfinden.

³⁷ Selektiv, da die Geschichte sich durch die Auswahl der präsentierten Geschichteseinzelteile zusammensetzt.

³⁸ Die Änderung des Blickwinkels (point of view) ist ein solches interaktives Element: Der Zuhörer kann den Blick auf die Geschichte ändern, jedoch den Fortlauf der Geschichte nicht beeinflussen. Der Erzählfluss der Geschichte kann zwar geändert werden, die eigentliche Erzählung jedoch bleibt gleich. Dadurch wird allerdings nur ein geringer Grad an zusätzlicher Immersion erreicht.

te verwalten³⁹. Dieses Verfahren wird auch diskrete nicht-lineare Narration⁴⁰ (Poly-Lineare Geschichten) genannt. Die poly-linearen Geschichten sind vor allem im Bereich des Hypertext bekannt, vgl. [Var99]. Es werden zwei Ansätze unterschieden, vgl. [vHBB00]:

- **Branching:** Die nichtlineare Geschichte ist aus verschiedenen Zweigen aufgebaut, siehe Bild 3.6. In jeden Zweig der nichtlinearen Geschichte sind feste Verzweigungspunkte integriert. An diesen Punkten müssen Entscheidungen getroffen werden, welche den Partizipanten der Geschichte in einen neuen Zweig der Geschichte führen. Für jeden Zweig existiert eine vordefinierte Dramaturgie, welche der Benutzer bis zum nächsten Entscheidungspunkt durchlebt.

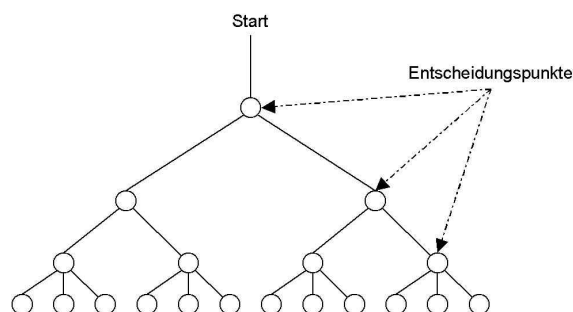


Fig. 3.6: Nichtlineare Erzählstruktur basierend auf Branching

- **String of Pearls:** Ähnlich zum Branching ist auch der String of Pearls aus verschiedenen Zweigen aufgebaut, siehe Bild 3.7. In jedem Zweig existieren *notwendige* Aufgaben und *zusätzliche* (quasi freiwillige) Aufgaben. Um von einem Zweig in den nächsten zu gelangen, müssen vom Benutzer alle notwendigen Aufgaben gelöst werden. Die zusätzlichen Aufgaben gestalten die Geschichte abwechslungsreicher. Der Benutzer muss nicht explizit wissen, welche der Aufgaben notwendig und welche zusätzlich sind.

Eine grobe⁴¹ Kontrolle wird erreicht, wenn die Reihenfolge der Plots nicht mehr in einer netzartigen Struktur angelegt sind. Eine solche Geschichte besteht nicht aus vorher definierten, linearen Versatzstücken. Tatsächlich wird durch eine Evaluationsfunktion die Publikumsaktion so in die Narration der Geschichte einbezogen, dass der nächste Plot aus der Summe der Publikumsaktionen, der Historie von Plots und dem zu erreichenden Ziel der Geschichte bestimmt wird. Die Narration bildet quasi eine

³⁹ Es gibt daher eine vordefinierte Szenen-Verzweigung, die Auswahl innerhalb der nächsten Szenen ist jedoch nicht eingeschränkt. Innerhalb einer Szene gibt es Raum für Variationen - je nach Reaktion des Publikums wird die nächste Szene eingeblendet.

⁴⁰ Durch gewisse Navigationsmechanismen kann der Zuschauer zwischen Sub-Geschichten mit linearer Narration auswählen. Hierbei kann der Benutzer den Fortlauf der Geschichte i.A. nicht ändern, er erhält aber eine benutzerspezifische Sichtweise auf die einzelnen Fragmente der Geschichte.

⁴¹ Das Wort grob bezieht sich auf die Granularität der Kontrolle, nicht auf die Art der Ausführung.

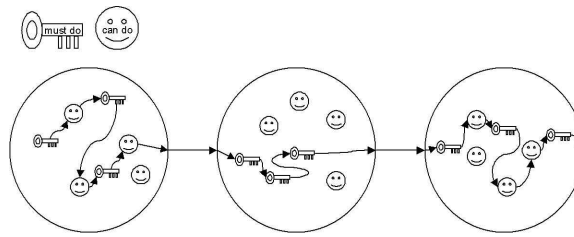


Fig. 3.7: Nichtlineare Erzählstruktur basierend auf String of Pearls

benutzerabhängige Funktion innerhalb eines relativen, nichtdiskreten Narrationsraumes⁴², siehe Bild 3.8.

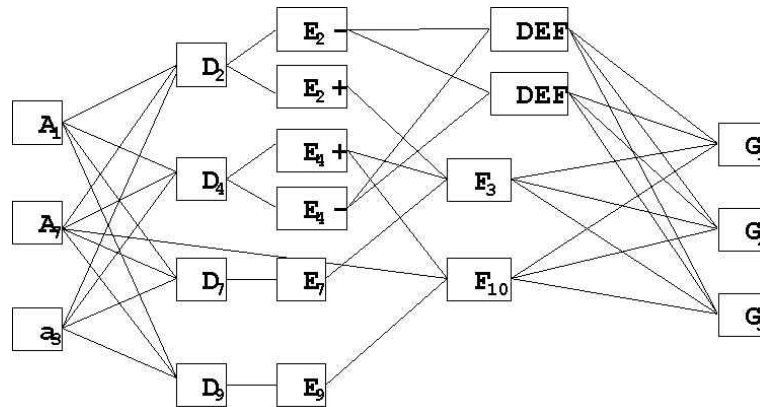


Fig. 3.8: Nichtlineare Erzählstruktur basierend auf morphologischen Funktionen

Generierung von Geschichten Ein nichtlinearers Geschichten-System kann die Geschichte mehr oder minder generativ darbieten. Das Spektrum reicht von starren, linearen Geschichten über eine Menge von Geschichten (Branching), bis hin zur völligen Neugenerierung einer Geschichtsstruktur.

Ist sich der Autor im klaren, welche Einflussgrößen den Verlauf der Geschichte und damit die Dramaturgie und auch die Immersion des Benutzers bestimmen, so gilt es ein entsprechendes dramaturgisches Verfahren⁴³ zur Modellierung der Geschichte zu wählen.

Je nach Ausprägung der Einflussgrößen einer nichtlinearen Geschichte wird ein anderes Maß an automatisierter erzählerischer Organisation benötigt. Der Begriff der *Narrative Intelligence*, im Deutschen mit *erzählerischer Intelligenz* oder mit *geschickter Erzählweise* übersetzt, wurde etwa 1990 durch die Narrative Intelligence Group am MIT begründet, vgl

⁴² Ein Beispiel für ein solches System ist das Oz-Projekt, siehe Kapitel 3.3.2.1 und auch der in Abbildung 3.8 gezeigte Ansatz von Propp, siehe 3.2.2.2.

⁴³ Diese dramaturgischen Verfahren sind speziell abhängig von der Art der nichtlinearen Geschichte und werden in Kapitel 3.2.2 beschrieben.

[MS99]. Die Ausrichtung der Forschung beschäftigte sich mit der Frage, wie Menschen mittels Erzählungen ihre Erfahrungen organisieren. Die Ursprünge der heterogenen, interdisziplinären Gruppe lagen in der KI - wurden aber um humanistische Einflüsse erweitert, um narrations-relevantes Wissen aus Gebieten verschiedenster Couleur wie Philosophie, Psychologie, Kognitionswissenschaft, Medientheorie, Literaturwissenschaft, künstliche Intelligenz (KI) und Mensch-Maschine-Interaktion zu diskutieren. Basierend auf den Einflussgrößen nichtlinearer Geschichten sowie der Kategorisierung von Mateas und Sengers, vgl. [MS99], werden in der vorliegenden Arbeit die daraus resultierenden Forschungen in die folgenden Bereiche unterteilt:

- Geschichten-generische Systeme

Der Schwerpunkt dieses Bereiches liegt in der Generierung von Geschichten. Generell wird zwischen Interaktiver Fiktion (lokale Kontrolle) und Interaktive Drama (globale Kontrolle) unterschieden.

Interactive Fiction: Zur Generierung werden häufig autonome Agenten - nach dem Prinzip des Emergent Narrative - genutzt. Diese sind Agenten, welche durch eine gewisse Autonomie eine Geschichte mittels der Interaktion, welche von ihnen ausgeht, aufspannen. Beispiele hierfür sind in dem Oz-System, vgl. [Mat97], oder im Erasmatazz System, vgl. [Cra99], zu finden. Dort interagiert der Benutzer mit agentengesteuerten Charakteren und Umgebungen. Aus dieser Interaktion heraus erzeugt sich eine Geschichte. Oftmals sind diese Geschichten wenig zielgerichtet, d.h. ohne wirkliche Handlungshöhepunkte. Der Ansatz ist in Grundzügen eine Simulation der Welt - der Zuschauer wird nicht von Höhepunkt zu Höhepunkt geführt, sondern muss jede Einzelheit der Welt miterleben - was den geringen Suspense des Zuschauers/Benutzers erklärt.

Interactive Drama verfolgt das Ziel, eine generierte Dramaturgie interaktiv darzubieten. Dieses Konzept widerspricht in weiten Teilen dem Gedanken der autonomen Agenten: Wenn in die Dramaturgie einer Geschichte eingegriffen wird, welche von autonomen Agenten 'produziert' wird, so muss auch in die Autonomie der Agenten eingegriffen werden - diese quasi zwangsgesteuert werden. Dadurch werden diese Systeme jedoch ihres Vorteils beraubt - die autonome Generierung von Verhalten. Aus diesem Grund sind die Konzepte zur Zwangssteuerung der Agenten selten komplexer als die Vorgabe bestimmter, einfacher Ziele und Motivationen. Zumeist erweisen sich diese dramatischen Ansätze jedoch nicht als umfassend genug, um tatsächlich mehr als eine Kurzgeschichte oder Szene einer Geschichte zu erzählen⁴⁴.

- Geschichten-erzählende Systeme

In diesem Bereich wird eine bereits existierende Geschichte interaktiv erzählt. Dabei lassen sich Autoren-basierte Systeme (diese Versuchen die Prozesse zu simulieren,

⁴⁴ Dies wird zum Beispiel bei den entsprechenden Erweiterungen des Oz-Systems, vgl. [Wey97], des Erasmatazz-Systems, vgl. [Cra01], oder des auf dem aristotelischen Drama aufbauenden System von Sgouros, vgl. [SPT97], deutlich.

welche in menschlichen Autoren ablaufen), geschichten-basierte Systeme (welche versuchen, mittels einer abstrakt-symbolischen Repräsentation Geschichten strukturiert zu vermitteln, zum Beispiel der Ansatz von Propp) und welt-basierte Systeme, welche ähnlich dem Prinzip des Emergent Narrative arbeiten.

- Narrative (erzählerische) Interfaces

In diesen Systemen wird die Narration zur Gestaltung der Mensch-Maschine Schnittstelle herangezogen. Das narrative Element bezieht sich dabei auf die Ästhetik⁴⁵ der Schnittstelle.

- Erzählstrukturen zur automatischen Wissensstrukturierung und Geschichten-Datenbanken.

Ansatz der KI, um eine effiziente Wissensstrukturierung zu ermöglichen, z.B. beim Einsatz von autonomen Agenten. Der Ansatz ist eng verwandt mit Datenbanken, welche Geschichten zum Strukturieren von Daten nutzen.

Geschichten-verstehende Systeme, welche das menschliche Verständnis einer Geschichte modellieren, werden explizit in der Kategorisierung nicht aufgeführt, da diese für die Thematik des Geschichtenerzählens nicht relevant sind. Der Schwerpunkt dieser Arbeit liegt auf dem interaktiven Erzählen, nicht auf dem Generieren von Geschichten. Im folgenden Kapitel wird auf die Geschichten-erzählenden Systeme eingegangen und diese detailliert diskutiert.

3.2.2 Automatisierte Geschichtsnarration

Um eine Geschichte automatisch und interaktiv erzählen zu lassen ist die Auswahl eines dramaturgischen Verfahrens unabdingbar, wenn der Suspense und die Immersion des Benutzers zwingend gewährleistet sein sollen. Je nach Verfahren müssen andere Daten vorgehalten werden, um die Erzählung der Geschichte zu realisieren.

Oftmals wird Dramaturgie mit der Intelligenz (im Sinne von KI) der virtuellen Akteure einer Geschichte gleichgesetzt⁴⁶. Nun ist die Intelligenz der einzelnen Charaktere einer Ge-

⁴⁵ So ist beim System Hypercafe, siehe Kapitel 3.3.1.2, bewusst eine Kaffeehausumgebung als Schnittstelle zum Benutzer gewählt, um ihm die mit einem Kaffeehaus verbundenen Verhaltensweisen ohne weitere Erläuterung als Bedienmetaphorik klar zu machen.

⁴⁶ Selbst wenn die Intelligenz der virtuellen Akteure gleichgesetzt ist mit der Intelligenz eines Menschen - d.h. der Akteur besitzt die Fähigkeit seine eigenen Ziele frei zu planen, er kann mit den Mitakteuren frei interagieren um diese Ziele zu erreichen - so reicht dies zur dramatischen Narration von Geschichten nicht aus. Es führt zwar zu einem regen Treiben auf der Bühne, Akteure interagieren miteinander, verfolgen ihre Ziele. Jedoch sind diese Ziele im allgemeinen wenig dramatisch und für den Zuschauer oder Mitspieler eher uninteressant anzusehen. Hierzu ein Beispiel bezüglich Suspense: Betrachtet man den folgenden Spezialfall mit den beiden Akteuren Julius und Julia: Julius möchte Julia zum Essen ausführen und kauft ihr einen Strauß Blumen in einem Blumenladen. Ist der Weg, den Julius vom Blumenladen zu Julia nimmt, interessant für die Geschichte? In den meisten Fällen wohl nicht - jedoch kann ohne ein Wissen um dramaturgische Entwicklungen und Höhepunkte das ganze lange Erleben des Weges durch das Publikum wohl nicht verhindert werden. Aus diesem Grund werden virtuelle Akteure oftmals separat modelliert. So sind Charakter- und Verhaltensmodelle die Grundlage für die Modellierung von Akteuren, welche sich im Rahmen einer Geschichte relativ frei auf der 'Bühne' der Geschichte bewegen, siehe [Woo00].

schichte sicherlich eine die Narration der Geschichte stark beeinflussende Komponente - jedoch können sie das Modell einer dramatischen Entwicklung nicht ersetzen.

Werkzeuge⁴⁷ zum Ausführen von Geschichten bewegen sich auf einer anderen Ebene - sie beschreiben szenische Handlung, den Verlauf einer Geschichte oder spezielle Höhepunkte einer Geschichte. Die verschiedenen Methoden und Ebenen des Geschichtenbeschreibens werden im Folgenden diskutiert und die entsprechenden Werkzeuge in einem Zusammenhang mit der Ebene der Geschichtenbeschreibung benannt.

3.2.2.1 Geschichtenbeschreibung auf Szenen-Ebene

Die Geschichtenbeschreibung auf Szenen-Ebene blickt auf eine lange Tradition in Computer Games zurück. Exemplarisch für diese Tradition steht die Roguelike Game Engine. Die Roguelike Game Engine stellt die Regeln für ein Spiel (eine interaktive, nichtlineare Geschichte), in einer amodalen Art dar. Die amodalen Regeln werden geteilt in Aktionsbeschreibungen und Charakter/Szenenbeschreibungen. Anhand einer Kombination der Regeln können komplexe Aktionen konsistent beschrieben werden. So ist eine Aktionsregel von der Art

-character: [-characteraction, -characterenter, -characterattack].

Eine Szenenbeschreibung ist von der Art

[time 0] Kitchen. A table. A man enters.

Anhand dieser beiden Regeln ist leicht zu erkennen, dass die Roguelike Game Engine tatsächlich bei den Aktionsbeschreibungen auf eine Ausformulierung des Regel-Atoms Character verzichtet. Der Character kann in Szenen-Regeln als Mann, Frau oder auch als Feuer dargestellt werden, je nach Sinn der Geschichte. Die Art des Geschichtsmanagements ist in dieser Form eher global, d.h. der globale Werdegang der Geschichte wird direkt durch die Regeln bestimmt. Tatsächlich existieren keine Konzepte, um die Dramaturgie über die gesamte Geschichte zu beschreiben.

3.2.2.2 Geschichtenbeschreibung auf Story-Ebene

Die Idee, ein Modell für eine bestimmte Kategorie von Geschichten zu finden, ist bekannt aus der Literatur, beispielsweise in der Klassifikation von wiederkehrenden Elementen der Handlung, vgl. [Goe94]. Diese Elemente eignen sich, um durch Umgruppierung neue Geschichten zu erzielen. Leider sind sie zu schwach, um eine wirkliche funktionale Einordnung der Handlungselemente zu erzielen. Das Modell von Propp⁴⁸ dagegen ist umfassend genug, um den Effekt der Neugestaltung von Geschichten aufgrund von Publikumsinteraktion zu gewähren.

⁴⁷ So sind zum Beispiel die Roguelike Game Engine, vgl. [rog00a], [rog00b], Dr. K— Project, vgl. [Ric99], Erasmus, vgl. [Cra00], oder Oz, vgl. [BWK92], Werkzeuge zum Beschreiben von Geschichten.

⁴⁸ Vorgestellt in Kapitel 3.1.1.

Propp's Modell wurde für Volksmärchen aufgestellt, entspricht aber in überraschender Weise dem gängigen Schema von 'großen' Hollywood-Produktionen⁴⁹. Das Regelwerk erlaubt eine Umstellung des Zusammenhangs von Szenen aufgrund einer morphologischen Klassifikation der Funktion von Szenen und Akteuren⁵⁰. Dadurch ist ein interaktiv erweiterbares, spannendes und immersives Geschichten-Modell gegeben.

3.2.2.3 Geschichtenbeschreibung auf Welt-Ebene

Ein gegensätzlicher Ansatz findet sich in Crawfords [Cra99] Erasmatron bzw. Erasmus. Dort werden die Regelwerke äußerst Komplex als Weltsimulation definiert. So besitzt Erasmus ein Modell des menschlichen Seins, das unter anderem menschliche Formen von Konversation, Verhandeln (in Form einer Microökonomie), eine Persönlichkeit (bestehend aus über 30 Charakterzügen, mindestens vier Launen und acht Dimensionen von Partnerschaft) und eine generelle Theorie von Aktionen-Reaktionen unterstützt. Dieser Ansatz ist eher orientiert an einer agentenbasierten Geschichte, die sich durch Mikrostrukturen wie Agentenpersönlichkeit aufbaut und durch die Kommunikation zwischen Agenten ihre Narration erfährt. So wird auf der unteren Abstraktionsebene durch den Erasmatron ein realistisches Verhalten von Charakteren bewirkt - etwa durch das Einhalten von territorialen Aspekten wie der Präsenz auf Bühnen. Soll das Verhalten eines Charakters jedoch auf einer kognitiven Ebene zielgerichtet proaktiv oder reaktiv sein, so muss dies vom Autoren der Welt durch entsprechende Verben vordefiniert werden. Durch diesen *Verbenansatz* kann, durch Definition von Rollen wie Subjekt und Objekt eines Verbs, komplexes Verhalten generiert werden. Zum Beispiel wird für das Verb *verstecken* die Rolle des Verfolgers definiert. Jeder Charakter kann in eine solche Rolle schlüpfen, er muss jedoch sinnvolle Reaktionsweisen (definiert wiederum über Verben) dafür aufweisen.

Mit diesem relativ einfachen Konzept können komplexe Welten definiert werden - dies geht über Reaktionen, Planen, Planausführungen über Berechnung von wahrscheinlichen Reaktionen auf erwartete Handlungen bis hin zu den Fähigkeiten zu Lügen, Geheimnisse verraten oder konspirativ handeln. Eine komplexe Dramaturgie kann damit nur sehr schwer verwirklicht werden, obwohl ein entsprechendes Verb PlotPoint nachträglich von Crawford eingefügt wurde, vgl. [Cra01].

⁴⁹ Auch exotische Genres können mit dem grundsätzlichen Modell der morphologischen Funktionen abgedeckt werden

⁵⁰ Die Konzeption und Implementierung eines Story Modells und einer Story Engine (als Story Engine wird ein Software System bezeichnet, welches ein generisches oder Geschichten-erzählendes System ist.) basierend auf Propp wurde vom Autoren der vorliegenden Arbeit in [BG01] beschrieben und wird in Kapitel 5.3 diskutiert. Weitere Arbeiten sind u.a. von Paiva, vgl. [PMP01], vorgenommen, welche mittels des Dramatis Personae Modells von Propp intelligente Agenten Geschichten erzählen lässt, welche in Grundzügen von Kindern vorgeformt sind.

3.2.2.4 Kombinierte Geschichtenbeschreibung

Als weiteres Beispiel einer nichtlinearen Erzählweise von Geschichten ist die Methodik von Oz⁵¹ zu nennen. Oz kombiniert die beiden Ansätze des globalen Kontrollings einer Geschichte und der lokalen Interaktion zwischen Agenten, welche Aufgrund ihrer Persönlichkeit die Narration einer Geschichte vorantreiben und eine Dramaturgie sowie Spannung aufbauen. Das Verhalten der Agenten wird von Oz ständig kontrolliert und je nach Publikumsinteraktion in eine gewisse Narrationsrichtung gelenkt. Die Oz-Engine ist nach dem Prinzip eines Schachcomputers aufgebaut: Jede Spielerinteraktion wird als ein Zug gewertet, der sich potentiell gegen das Erreichen eines Geschichtsziels auswirken könnte. Die Systemantwort berechnet sich aus der Evaluation der bisherigen Historie der Geschichte sowie aller möglichen Antworten, insbesondere deren Auswirkung auf die Möglichkeiten des Publikums, sich noch weiter vom Geschichtsziel zu entfernen bzw. sich dem Geschichtsziel wieder anzunähern.

3.2.2.5 Zusammenfassung

In der Spieleindustrie finden sich noch sehr viele weitere Story Engines, die sich jedoch nicht wesentlich von den oben gezeigten Ansätzen unterscheiden. Tatsächlich ist der größte Anteil von Spiele-Story Engines untrennbar mit der Rendering-Engine der entsprechenden Spiele kombiniert - was eine Portierung dieser Engines auf neue Spiele fast unmöglich scheinen lässt. Jedoch ist die Trennung von Erzählung und Visualisierung auch in der Spieleindustrie bei neuen Produkten erkennbar - auch wenn die Spieleindustrie im allgemeinen eher auf eine Verbesserung der AI von Spielfiguren denn auf eine Verbesserungen der geschichtserzählenden Elemente im allgemeinen abzielt.

Für den Einsatz von Video-Clips gilt zu bedenken, dass der Video-Clip - als Szene interpretiert - keine nachträgliche Dramaturgie auf Szenen-Ebene zulässt - so empfiehlt sich für die Narration auf Basis von Video ein Story-Ebenen Ansatz wie der von Propp. Im Gegensatz zum aktuellen Trend der Spieleindustrie wird das Hauptaugenmerk nicht auf eine besonders starke Spiele-Intelligenz der einzelnen Charaktere gelegt, sondern eine zusammenhängende Erzählweise der Geschichte mit einer starken inhaltsbezogenen Intelligenz der Figuren favorisiert.

3.2.3 Funktion von interaktiv-erzählerischen Elementen

Die Inhaltsnarration einer Geschichte bestimmt, welche narrative (Story-) Situation an der entsprechenden Stelle zwischen Geschichte und Publikum angestrebt wird.

Interaktiv-erzählerische Elemente orientieren sich an dem Begriff der *polymorphic beats*, welcher durch Mateas und Sengers [MS00] eingeführt wurde. Ein Beat gilt bei Mateas und Stern als atomarer Bestandteil der Geschichte, d.h. der kleinste Wert, dessen Änderung eine Änderung der Geschichte provoziert - je nach Interaktion des Benutzers (mit durchaus unterschiedlichen Möglichkeiten des Ergebnisses) mit anderen Auswirkungen auf die

⁵¹ Oz ist in Kapitel 3.3.2.1 detailliert beschrieben

Geschichte.

Die Story-Situationen können grundsätzlich nach ihrer Funktion für eine Geschichte unterschieden werden. So werden Story-Situationen auf Szenen-Ebene oft über Speech-Acts, vgl. [Gri75], definiert:

- **Kommunikations-Status:** In dieser Menge werden Speech-Acts zusammengefasst, welche das Publikum erkennen lassen, ob und wie das geschichtenerzählende System - zum Beispiel anhand von virtuellen Actoren - auf Benutzer-Interaktion reagiert. Dazu gehören zum Beispiel das Zeigen einer nicht-interpretierbaren⁵² Situation, das Abbrechen einer Kommunikations-Schleife oder das Zeigen von 'Leerlaufverhalten'.
- **Kommunikations-Management:** Neben dem Anzeigen des reinen Verständnisses sollte die Kommunikation mit dem Benutzer in gewissen Bahnen geregelt werden. Hierzu gehören zum Beispiel folgende Speech Acts:
 - Beschränkung des Publikums auf die möglichen Interaktionen.
 - Turntaking⁵³-Aufforderung, d.h. das Angebot an den Benutzer an dieser Stelle den aktiven Part einer Kommunikation zwischen System (Geschichte) und Publikum einzunehmen bzw. die Aufforderung, den aktiven Part der Kommunikation an die Geschichte abzugeben.
- **Aktionsaufforderung,** die sich auf vorangegangene Informationen bezieht.
- **Aktionsaufforderung,** die sich auf zukünftige Informationen bezieht.
- **Reaktion auf Publikumsintervention,** die sich keiner vorherigen Kommunikationssituation zuordnen lässt.
- **Feedback-Aktionen** wie zum Beispiel die Reaktion auf eine Navigations-Aktion des Publikums.
- **Emotions-Management:** Der bewusste Versuch, durch das Zeigen einer der Situation angepassten Emotion, unter Berücksichtigung der Publikumsemotionen, eine emotionale Situation für das Publikum nachvollziehbar zu machen und die Emotion des Publikums in gewisse Bahnen zu lenken.

Auf Geschichte-Ebene dienen Story-Situationen⁵⁴ dazu, den weiteren Verlauf der Geschichte anzuzeigen und - durch Auswahl einer entsprechenden Story-Situation durch den Benutzer - den weiteren Verlauf der Geschichte zu bestimmen.

⁵² Nicht-interpretierbar bedeutet in diesem Fall eine für das System nicht eindeutige Situation.

⁵³ Der englische Begriff des Turn taking wird in dieser Arbeit (eingedeutscht) als Turntaking notiert.

⁵⁴ Eine Story Situation, auch genannt Story-Act, ist speziell für die einzelne Geschichte zu definieren und entzieht sich einer allgemeinen Definition, siehe Kapitel 5.3.1.

3.2.4 Literaturtheoretische Analyse

Interaktives Geschichtenerzählen ist in literaturtheoretischer Hinsicht ein Nachfolger des literarischen Genres der Computer Literatur, siehe Lorenz [Lor92], und findet seine Ursprünge in der automatisierten Generierung von Geschichten. Geht man jedoch davon aus, dass die interaktiven Geschichten durch einen Autoren vordefiniert werden, so scheint es sinnvoll, interaktives Geschichtenerzählen in seinem Zusammenhang mit verschiedenen Literaturkategorien und Genres zu definieren, vgl. Braun et al [BSH02]. Im folgenden werden die Grundlagen des interaktiven Geschichtenerzählens im Bezug auf die literarische Kategorie, narrative Elemente und die dramatische Klassifikation von Literatur notiert.

3.2.4.1 Literarische Kategorie

Für die deutsche Sprache werden von Goethe [Goe60, 126-267] drei literarische Gattungen definiert - die Lyrik, die Epik und die Dramatik. Die Einordnung von interaktivem Geschichtenerzählen in diese Kategorien ist problematisch:

- Interaktive Geschichten besitzen einen Erzähler, dies weist auf die Kategorie Epic hin.
- Die mimetischen Präsentationsweisen beim interaktiven Geschichtenerzählen und die vordefinierte Konstruktion von Geschichten scheinen jedoch von der Kategorie Dramatik zu sein.

So ergibt es sich, dass interaktive Geschichten als Epen mit dramatischen Elementen gesehen werden müssen, welche durch interaktive Elemente unterstützt werden. Der Spagat zwischen Epik und Dramatik ist unter anderem im Genre der Novelle für die Epik und im epischen Theater (vgl. Brecht [Bre63]) der Dramatik vollzogen.

Die Novelle ist in der Gattung Epik, jedoch ausgezeichnet mit einer internen dramatischen Struktur: Die Novelle zeigt Ereignisse, welche als Wendepunkte im Leben der Protagonisten angesehen werden. Diese Ereignisse provozieren Aktionen, welche sich zu einem Höhepunkt auswachsen. Die Novelle bietet somit die dramatische Struktur einer Exposition, eines Höhepunkt, eines retardierenden Moments und eines Endes, wie von Braak [Bra69, pp.550] notiert. Die folgende literaturtheoretische Definition ist somit adäquat für interaktives Geschichtenerzählen:

Interaktives Geschichtenerzählen ist die interaktive, mimetische Präsentation einer Novelle

Brecht's Episches Theater kommt aus der entgegengesetzten Richtung - Brecht verbindet Szenen ohne dramatisches Konstrukt, sie stehen gleichwertig Seite an Seite. Dies provoziert eine epische Form des Dramas. Brecht befreit das Publikum damit von seiner passiven Rolle des Zuschauers und versucht, eine tiefe Auseinandersetzung des Publikums mit dem

dargebotenen Material zu ermöglichen. Natürlich steht die Auflösung des dramatischen Konstrukts im direkten Gegensatz zu der Dramaturgie, welche das interaktive Geschichtenerzählen erreichen möchte. Insofern ist der Ansatz von Brecht für das interaktive Geschichtenerzählen nicht sinngebend.

3.2.4.2 Narrative Elemente

Die narrativen Elemente der Literatur sind von Weber [Web98] definiert:

- Narration ist das serielle Adressieren von temporal spezifizierten Ereignissen.
- Narration bezieht sich auf das nicht-aktuelle.
- Erzähler befinden sich außerhalb der Erzählung.
- Narration besitzt zwei generelle Orientierungspunkte - die erzählte Zeit (Ich-Hier-Jetzt-System der Protagonisten) und die Erzählzeit (Ich-Hier-Jetzt-System des Erzählers).
- Die Erzählung spricht immer ein Publikum an.
- Narration ist hintereinander ab folgend (sukzessiv) und nicht perfektisch (zeigt die Aktion, nicht nur das Ergebnis).

Diese Kriterien gelten ebenfalls für das interaktive Geschichtenerzählen, jedoch sind die ersten beiden Kriterien eingeschränkt:

Interaktives Geschichtenerzählen bezieht sich ebenfalls auf das serielle Adressieren von temporalen Umständen, diese werden jedoch nichtlinear und interaktiv im Zusammenspiel mit dem Publikum zusammen gestellt. Aus diesem Grunde sind die einzelnen temporalen Punkte nicht vordefiniert, ebenso sind die Elemente der Geschichte nicht als dem Publikum vorab bekannte Sequenz definiert.

Interaktives Geschichtenerzählen impliziert eine aktuelle Erfahrung des Publikums und bezieht sich somit nicht auf nicht-aktuelles, sondern auf aktuelle, quasi gerade stattfindende Ereignisse. Dies ist als weiterer Hinweis auf die dramatische Struktur des interaktiven, nicht linearen Geschichtenerzählens zu sehen, da ein Drama ebenfalls auf aktuellen Ereignissen basiert. Wilder nennt dies: “On the stage it is always now”, siehe [Web98, pp.24].

3.2.4.3 Dramatische Klassifikation

Zur Diskussion der dramatischen Elemente des interaktiven Geschichtenerzählens wird auf Staiger’s Formenkategorien verwiesen, siehe [Bra69, p.117]. Staiger definiert unter anderem die Formkategorie des Dramatischen. Der Ausdruck *Drama* beschreibt bei Staiger das spielen von Charakteren, d.h. eine memetische Präsentation. Braak konkretisiert als Schlüssel zur Dramatik den Dialog, falls der Dialog Spannung erzeugt, siehe [Web98, pp.117]. Die

Aktionen in einem Drama werden zu einem Höhepunkt getrieben, aus dieser Direktionalität heraus ergibt sich eine Spannung beim Publikum, welches den Höhepunkt erwartet. Interaktives Geschichtenerzählen adaptiert eben diese Charakteristik: Aktionen werden ausschließlich dann präsentiert, wenn die Präsentation der Aktionen relevant und wichtig für den weiteren Verlauf der Geschichte ist.

Innerhalb der Formkategorie des Dramatischen werden verschiedene Sub-Kategorien unterschieden. So finden sich Einortdrama und Bewegungsdrama. Auf den ersten Blick scheint Interaktives Geschichtenerzählen eine Art Bewegungsdrama zu sein, da das Publikum sich zumeist frei in der Szenerie der Geschichte bewegen kann. Da das Publikum jedoch zu einer Zeit nur an einem Ort sein kann, beim interaktiven Geschichtenerzählen jedoch eventuell an anderen Plätzen ebenfalls relevante Dinge passieren, welche das Publikum nicht sieht, muß das Publikum schließlich selbst als dramatisch handelnde Person (mit eingeschränktem, subjektivem Wissen) angesehen werden. Aus diesem Argument ergibt sich die Nutzung von Einortdrama-Techniken für interaktives Geschichtenerzählen.

Weiterhin wird interaktives Geschichtenerzählen als geschlossenes Drama angesehen. Ein geschlossenes Drama gibt einen logischen Ausschnitt einer geschlossenen Umwelt in einer fiktiven Realität zum besten. Aktionen, Raum und Zeit sind konsistent nach Art einer gut argumentierten Geschichte des Dramatica-Schemas (vgl. [Dra]). Im Gegensatz dazu steht das offene Drama, welches eine Vielzahl von dramatisch gleichwertigen Aktionen, Räumen und Zeitlinien bietet, siehe Klotz [Bra69, p.225]. Auch wenn die szenische Anordnung einer interaktiven Geschichte an das offene Drama erinnert, so sind die Szenen selbst doch voneinander abhängig (d.h. abhängig von der gesamten Geschichte), was einem geschlossenen Drama entspricht.

Zusammenfassend kann die Charakteristik einer interaktiven Geschichte in der Formkategorie des Dramatischen als geschlossenes Einortdrama angegeben werden.

3.2.5 Zusammenfassung

Werden die in der Einleitung beschriebenen Voraussetzungen in Betracht gezogen so wird klar, dass ein einzelnes Konzept der nichtlinearen Narration nicht ausreicht, um tatsächlich interaktive Narration in ihrer ganzen Komplexität mit Video zu beschreiben, da durch die szenische Festlegung von Video-Clips das Erreichen der szenischen Funktionen von interaktiv erzählerischen Elementen⁵⁵ nicht gelingt. An dieser Stelle wird ein kombinierter Ansatz der Story- und Szenebeschreibung gefordert, um eine tatsächliche Dramaturgie der Narration mit einer Berücksichtigung der elementaren szenischen Ansprüche des Benutzers zu gewährleisten und Interaktives Geschichtenerzählen im literaturtheoretischen Sinne zu ermöglichen. Dieser Aspekt wird in Kapitel 3.4 aufgegriffen und in den entsprechenden generellen Anforderungskatalog dieser Arbeit integriert.

⁵⁵ Vgl. Kapitel 3.2.3.

3.3 Aktuelle Ansätze und Konzepte

Videopräsentationen wurden bisher und werden zur Zeit im Wesentlichen für eine lineare Präsentation zusammengestellt. Aus diesem Grund wurden überwiegend polylineare Narrationsmethoden für nichtlineare Videopräsentationen entwickelt. Die Narration der Geschichten ist nicht wirklich frei programmierbar - sondern immer auf die Video-Clips eingeschränkt, die zur Bildung der Narration zur Verfügung stehen. Der Clip-basierte Ansatz von digitalem Video basiert somit auf der Selektion eines zur aktuellen Gesichtssituation entsprechend geeigneten weiteren Clips. Im Clip selbst sind sowohl Charaktere als auch Fortgang der Geschichte enthalten. Diese Narrationsform kann als *selektiv* bezeichnet werden.

Im Gegensatz dazu stehen VR-Systeme, die zwar kein echtes, menschliches Schauspiel darstellen können, jedoch i.A. die Möglichkeit besitzen, aus verschiedenen primitiven Animationen von Avatar-Verhalten eine komplexe Verhaltensdarstellung zu entwickeln, was als *generativ* bezeichnet wird.

Obwohl die Verhaltensbibliothek eines Agenten beschränkt ist, kann auf Grund der geringen Granularität der einzelnen Verhaltensbausteine doch ein komplexeres Verhalten generiert werden als dies mit Clips (in welchen das Verhalten von Spielfiguren explizit abgelegt ist) möglich ist. Das selbe gilt für Emotion, welche u.a. durch Verhalten gezeigt werden. Emotion und Verhalten können im Fall von generierenden Systemen wesentlich besser miteinander kombiniert werden. Clip-basierte Systeme müssen zum Erreichen ähnlicher Ziele vollständig im Bezug auf Verhalten und Emotion der Spielfiguren indexiert werden.

3.3.1 Digital Video-basierte Ansätze

Im Folgenden werden richtungsweisende Systeme beschrieben, die sich mit nichtlinearer Videopräsentation auseinandersetzen. Dabei wird zwischen generischen Systemen ohne spezielle Ästhetik und Systemen mit einer speziellen Ästhetik unterschieden.

3.3.1.1 Generische Ansätze

Im Bereich des streaming Video- und Audio ist die seit 1995 in diesem Gebiet tätige Firma RealNetworks, vgl. [HG97], Marktführer. Das System der Firma hat seit Ende 1998 die Versionbezeichnung G2 und unterteilt sich in eine Präsentationskomponente (RealPlayer), eine Serverkomponente (RealServer) und ein Autorensystem (RealProducer). Das RealSystem nutzt SMIL⁵⁶, vgl. [W3C98].

Die Interaktionsmöglichkeiten des RealSystem sind auf sogenannte Videorecorderfunktionalität und Annotationen des Videos (Hyperlinks)⁵⁷ beschränkt.

⁵⁶ Synchronized Multimedia Integration Language, ein Standard des W3C (World Wide Web Consortium), um Multimediapräsentationen zu beschreiben.

⁵⁷ Diese Begrifflichkeiten sind in Kapitel 4.3.2 definiert.

Real Media kennt zwei Arten von Hyperlinks: selbstauslösende Hyperlinks, bei Real Ereignisse genannt, und Image-Maps, d.h. anklickbare Bereiche innerhalb eines Videos.

- Ereignisse haben einen Start- und Endzeitpunkt, den Namen des Frames, in dem sie erscheinen sollen, und eine URL. Ereignisse können prinzipiell nur auf Dokumente, die mittels http übertragen werden, verweisen. Während ein Clip abgespielt wird, werden zum jeweiligen Startzeitpunkt die Ereignisse ausgelöst, d.h. die URL wird im Standard-Browser geöffnet. Das Öffnen eines Clips oder der Sprung innerhalb des aktuellen Clips ist über Ereignisse nicht möglich.
- Image-Maps haben eine Form (Rechteck, Kreis oder Polygonzug), einen Start- und Endzeitpunkt, einen Aktionstyp und eine ALT-Text⁵⁸. Klickt der Benutzer in den Bereich des Image-Maps, so wird die zugeordnete Aktion ausgeführt. Der Benutzer bekommt kein visuelles Feedback, d.h. der Bereich des Image-Maps wird weder im Video markiert, noch weiß der Benutzer, wie lange der Bereich aktiv ist. Es werden die in Tabelle 3.1 genannten drei Aktionstypen unterschieden.

Ereignisse und Image-Maps werden mit einem Zusatzprogramm zum RealProducer, dem RMEvent-Editor, in eine RealMedia-Datei eingebettet. Dabei wird dem RMEvent-Editor jeweils eine Textdatei für die Ereignisse und eine für die Image-Maps übergeben.

Tab. 3.1: Die Aktionstypen für Image-Maps in RealSystem

Aktionstyp	Beispiel	Bedeutung
PLAYER	PLAYER= "pnm://video.real.com/ welcome.rm"	Gibt einen neuen Clip an, der im RealPlayer gestartet werden soll.
URL	URL= "http://www.real.com"	Gibt eine URL an, die im Browser angezeigt werden soll.
SEEK	SEEK=0:0:0:5:0	Gibt eine Zeit im aktuellen Clip an, zu der gesprungen werden soll.

Zusätzlich besteht die Möglichkeit mittels SMIL ebenfalls ImageMaps⁵⁹ zu definieren - sie werden im SMIL Protokoll jedoch anchor⁶⁰ genannt.

SMIL wurde im Juni 1998 vom W3C verabschiedet. SMIL ist kein Binärformat und kann deshalb direkt durch einen Autoren geschrieben werden - es ähnelt in seinem Aufbau den von HTML her bekannten Strukturen. Durch die strikte Trennung von Medien und deren Zusammenspiel in einer Präsentation kann SMIL weitgehend medienunabhängig eingesetzt werden. SMIL bietet die folgenden Möglichkeiten:

⁵⁸ Der ALT-Text wird in der Statuszeile des RealPlayers angezeigt, wenn sich die Maus über dem Bereich des Image-Maps befindet.

⁵⁹ Gegenüber den Image-Maps des RMEvent-Editors kann die SMIL-Variante allerdings nur rechteckige Bereiche definieren. Die Angabe eines zusätzlichen Informationstextes, der beim Überfahren des Bereichs mit der Maus als Tooltip angezeigt werden kann, ist nicht vorgesehen.

⁶⁰ Ein anchor besteht aus einer URL, Koordinaten für ein Rechteck sowie Beginn- und Endzeit.

- Definition eines Layouts durch Regionen, denen dann einzelne Medien zugeordnet werden können.
- Paralleles oder sequentielles Abspielen von Medienclips, zugeordnet zu einer Region.
- Festlegen des Start- und Endzeitpunktes sowie der Dauer eines Mediums.
- Wiederholen eines Mediums.
- Prozedurale Struktur durch eine Switch-Anweisung.

Es ist denkbar, Präsentationen per Hand oder über ein mächtiges Tool zu ändern oder auch dem späteren Betrachter die Möglichkeit zu geben, die Präsentation an seine Bedürfnisse anzupassen - Untertitel zu einem Video könnten z.B. an- oder abgeschaltet oder in der Größe geändert werden. Bei der Definition des Layouts kann den einzelnen Regionen ein Tiefenindex gegeben werden, der bestimmt, in welcher Reihenfolge Objekte auf dem Bildschirm sichtbar werden. Alle weiteren SMIL Elemente können dann beliebig verschachtelt werden.

Ein Beispiel soll die Sprache weiter erläutern:

```
<smil> <head>
  <layout>
    <root-layout background-color="maroon" width="330"
                                   height="415"/>
    <region id="videoregion1" top="5" left="5" width="320"
                                   height="200" z-index="2"/>
    <region id="videoregion2" top="180" left="50" width="320"
                                   height="200" z-index="1"/>
  </layout>
</head>
  <body>
<par>

  <audio src="Level_1Audio.rm"/>
  <video src="level_1Video.rm"      region="videoregion1"/>
  <seq>
    <video src="nachrichten_Nr2.rm"  region="videoregion2"/>
    <video src="nachrichten.rm"      region="videoregion2"/>
  </seq>
  <audio src="pingmulti.rm" begin="10s" dur="6s"/>
  <audio src="pingmulti.rm" begin="20s" dur="4s"/>
  <audio src="pingmulti.rm" begin="30s" dur="2s"/>
</par>
</body>
</smil>
```

In diesem Beispiel werden zunächst eine Hintergrundfarbe und dann zwei sich überlappende Regionen definiert. Würde kein z-Index angegeben, so läge die zuletzt definierte Region über der ersten. Da aber der z-Index der Region *videoregion1* höher ist als der, der *videoregion2*, überlappt die *videoregion1* die *videoregion2*. Weiter ist zu sehen, dass die meisten Elemente parallel gespielt werden, wobei das Audioobjekt *pingmulti* zu verschiedenen Zeitpunkten mit unterschiedlichen Längen abgespielt wird. Nur das Video *nachrichten* wird erst nach Beendigung des Videos *nachrichten_Nr2* sequentiell abgespielt.

Mittels SMIL und mittels einer RealMedia internen Struktur sind synchron zum Video präsentierte, weitere Informationen und graphische Annotationen⁶¹ im Video möglich. Die nichtlineare Narration erfolgt dabei ähnlich einer Hypertextstory, allerdings erweitert um den Faktor Zeit. Die Ästhetik von RealMedia ist nicht spezifisch einem Szenario oder einer Geschichte angepasst, da RealMedia ein generischer Ansatz zur nichtlinearen Präsentation von Video ist und sich keiner speziellen Anwendung verpflichtet.

MPEG-4 (Motion Picture Expert Group Format 4, vgl. [Koe01]), ist die Weiterentwicklung der bestehenden ISO-Standards MPEG-1 und MPEG-2. Während MPEG-1 und MPEG-2 sich ausschließlich mit der effizienten Kompression von rechteckigen Videodaten befassen - MPEG-2 wird in vielen Set-Top-Boxen und auch bei der DVD verwendet - geht man bei MPEG-4 wesentlich weiter. Mit MPEG-4 werden nicht Videos kodiert, sondern komplexe Szenen beschrieben. Diese können sich aus den verschiedensten einfachen Medienobjekten zusammen setzen. Beispiele sind:

- Stehende Bilder (z.B. als fester Hintergrund)
- Videoobjekte (z.B. eine sprechende Person - mit oder ohne Hintergrund)
- Audioobjekte (z.B. eine Stimme, die der Person gehört)
- Texte
- Synthetische Köpfe und dazu assoziierter Text, aus dem die Sprachausgabe generiert und Informationen zur Animation des Kopfes geholt werden
- Synthetische Klänge

Ein kodiertes Objekt enthält sowohl beschreibende Informationen, die zur Visualisierung benötigt werden, als auch die zu streamenden Daten. Die Objekte sind jeweils unabhängig von anderen Objekten der Szene kodiert und können frei in der Szene platziert werden. MPEG-4 verfügt über die Möglichkeit der Interaktion⁶² auf den dargestellten Objekten. Um Objekte einer Szene anzusprechen, gibt es in MPEG-4 zwei verschiedene Multiplexer:

- Der Multiplexer FlexMux faßt elementare Objekte, die durch einen elementaren Stream (ES) repräsentiert werden, zusammen. Das Zusammenfassen kann hierbei nach logischen Objekten oder aber nach Quality of Service (QoS)-Anforderungen erfolgen.

⁶¹ In Form von Hyperlinks zur nichtlinearen Narration von Video.

⁶² Die Interaktionsmöglichkeit muss explizit durch die entsprechenden MPEG-4 Player implementiert sein.

- Für den zweiten Multiplexer, TransMux, wird lediglich die Schnittstelle durch MPEG-4 definiert. Die spezifische Umsetzung wird den für die jeweiligen Protokolle zuständigen Einrichtungen⁶³ überlassen.

Insgesamt ist MPEG-4 sehr generisch aufgebaut, entsprechende Player/Encoder existieren derzeit als teilimplementierte Lösungen.

Der sogenannte Media Player, vgl. [Obe99], von Microsoft bietet neben eigenen proprietären Audio- und Videoformaten auch MPEG-4 Version 3⁶⁴ an, die in einer Datei im Advanced Streaming Format⁶⁵ (ASF) abgelegt werden. Der MediaPlayer kann SMIL nicht interpretieren. Von Microsoft ist eine Beschreibungssprache namens Advanced Stream Redirector (ASX) vorgesehen, die zeitliche Abläufe steuert, diese kann Multimediaclips nacheinander, jedoch nicht parallel abspielen.

Eine sehr frühe Implementierung eines nichtlinearen Videopräsentationssystems ist Vosaic, vgl. [CTCL96]. Dieses System ermöglicht, ähnlich wie RealMedia, eine graphische Annotation von Video. Auch bei Vosaic lässt sich eine Art Hypertextstory mit Video realisieren. Ebenso wie RealMedia besitzt dieser Ansatz keine eigene Ästhetik und ist keiner speziellen Anwendung verpflichtet.

Zusammenfassend ist zu erkennen, dass die narrativen Elemente in den generischen Ansätzen von digitalem Video nicht modelliert werden. Die Granularität der Narrationskontrolle erfolgt auf einem Branching-Niveau, die Ansiedlung der Kontrolle ist Lokal, eine generische Bildung von Geschichten erfolgt nicht. Insofern erfüllen die generischen Digital Video basierten Ansätze nicht die minimalen Anforderungen an eine nichtlineare Narration, wie sie in Kapitel 3.2.5 gefordert werden.

3.3.1.2 Hypercafe

Das von Sawhney, Balcom und Smith entwickelte Hypercafe, vgl. [SBS96], ist ein experimenteller Hypervideo Prototyp (Hypervideo interpretiert als Kombination von digitalem Video und Hypertext), welcher einen navigatorischen Zugang zur nichtlinearen Geschichte gestattet. Das Szenario von Hypercafe ist dabei relativ beschränkt: Das Programm bringt das Publikum in ein virtuelles Cafe, welches aus Video-Clips besteht. Die einzelnen Clips zeigen die Konversation verschiedener Spielfiguren. Das Publikum, in diesem Fall die Einzahl, nämlich der Zuschauer, kann verschiedenen Konversationen folgen und mittels temporaler, spatial-temporaler und textueller Links alternative Narrationen ansehen.

⁶³ Für das Protokoll User Datagram Protocol/Internet Protocol UDP/IP wäre hier z.B. denkbar einfach eine Netzwerkverbindung pro Stream aufzubauen. Nach dem Transport der Daten über ein Netzwerk werden die Daten in der Delivery Multimedia Integration Framework (DMIF)-Schicht wieder demultiplexed, um in die Synchronisations-Schicht (SL, Sync Layer) weitergereicht zu werden. Dort werden die zeitlichen Abläufe synchronisiert und die elementaren Streams an die entsprechenden Dekoder weitergeleitet. Aus den Informationen über den Szenenaufbau und den dekodierten Objekten wird dann die Szene aufgebaut.

⁶⁴ Das Microsoft MPEG-4 Format unterstützt nur ein einziges mögliches Objekt - es nutzt nur Video-Funktionalitäten. Animierte Gesichter oder andere Objekte von MPEG-4 können mit dem Media On-Demand Producer nicht integriert werden.

⁶⁵ Der dazu benötigte Codec liegt dem Media On-Demand Producer bei.

Die Ästhetik eines Cafehauses ergibt sich dabei aus der Art, wie die Video-Clips präsentiert werden: Kontinuierlich spielen die Clips die Konversation an verschiedenen 'Tischen' (Video-Clips) des Cafehauses ab. Die Visite des Publikums im Cafehaus wird somit vergleichbar mit dem Echtzeit-Video des Lebens⁶⁶, welches in einem Cafehaus spielt.

Die Struktur von Hypercafe ist in 3 Ebenen geteilt:

- Auf der ersten Ebene wird eine Übersicht der Tische im Cafehaus gezeigt - die Kamera fährt von Tisch zu Tisch, das Publikum hat die Möglichkeit kurze Eindrücke der Konversation aufzuschnappen.
- Die zweite Ebene zeigt die Konversation an einem Tisch. Durch das Aufblenden von Video-Fenstern kann zwischen verschiedenen alternativen Verläufen einer Konversation ausgewählt werden.
- Die dritte Ebene bietet eine Übersicht der Gespräche am Tisch bzw. die Sicht vom Tisch auf andere Tische.

Dies impliziert eine 2-Ebenen Struktur des Cafes - einmal eine Sichtweise auf das gesamte Cafe, quasi als Gast, der zwischen den Tischen steht. Des Weiteren ist die Sicht eines am Tisch sitzenden Gastes möglich.

Die Navigation in Hypercafe erfolgt mittels Links:

- Temporale Links: Diese Links werden unterschieden in Videolinks und Textlinks.
 - Videolinks: Zum Ausgangsvideo werden temporär zusätzliche Video-Clips in temporalen Fenstern eingespielt. Diese zusätzlichen Video-Clips geben dem Publikum in einem begrenzten Zeitrahmen (3-5 Sekunden), eine abhängig von der narrativen Situation des Ausgangsvideos, mögliche alternative Narration des Videoinhaltes.
 - Textlinks: Parallel zu einem Video-Clip eingeblendete, assoziative Textlinks führen zu entsprechenden Video-Clips.
- Spatial-temporale Links: Innerhalb eines Video-Clips werden anklickbare Objekte durch eine Mauszeiger-Veränderung dargestellt. Da die meisten Objekte nur temporär sichtbar sind, ist das Auswählen des Objektes nur temporär möglich. Das Anklicken eines Objektes führt zu einem dem jeweiligen Objekt entsprechenden Video-Clip.

Hypercafe bietet ein entsprechendes Framework für das Erstellen und Präsentieren von Videos. Das Framework bietet die Zusammenstellung von nichtlinearen Videos über die Verknüpfung von Szenen untereinander, wobei durch das System zwischen einer linearen Sequenz und einem Link quasi kein Unterschied besteht. Dies macht es dem Autoren leicht, Sequenzen sowie die Verlinkung der Szenen untereinander zu handhaben. Die Szenen an sich können in verschiedenen Sequenzen genutzt werden, was für einen ähnlichen Effekt sorgt

⁶⁶ Tatsächlich eine sehr blumige Beschreibung der Autoren.

wie bei Joyce's *Afternoon, a Story*⁶⁷. Die Szenen erhalten, in unterschiedlicher Reihenfolge abgespielt, eine andere Bedeutung aufgrund des veränderten Kontexts. Da die Videos im wesentlichen Konversationen darstellen, ist eine entsprechende Produktion des Videos (Nahaufnahmen, Shoulder-Shots) unumgänglich.

Tatsächlich ist die Narration der Geschichte nichtlinear - der Plot der Geschichte ergibt sich mittels Branching⁶⁸. Der Zugang zur nichtlinearen Narration erfolgt in Hypercafe ausschließlich über graphische Links. Die Akustik des Videos wird, trotz des Schwerpunktes auf der Konversation von Spielfiguren nicht annotiert. Die Narration kann nicht konversational durch das Publikum beeinflusst werden - was nicht adequat ist für ein System, das im Wesentlichen an Konversation teilhaben lassen möchte.

Der ästhetische Ansatz von Hypercafe lässt die Folgerung zu, das Hypercafe kein Werkzeug zur Modellierung von Geschichten ist, sondern ein narratives Interface⁶⁹, welches die Narrationsmetapher des Kaffeehauses benutzt. Insofern erfüllen die Digital Video basierten Ansätze nicht die minimalen Anforderungen an eine nichtlineare Narration, wie sie in Kapitel 3.2.5 gefordert werden.

3.3.2 VR-basierte Ansätze

Die grundlegenden narrativen Einzelteile einer Geschichte werden in VR-basierten Ansätzen nicht explizit in Video-Clips gespeichert (ähnlich zum szenischen Begriff des Feature-Films), Vielmehr werden die Bestandteile der Narration generativ entwickelt über die Primitive der Animation. Die daraus folgenden Möglichkeiten erlauben eine Kontrolle der Narration einer Geschichte über die Lokation der Kontrollelemente, die Granularität der Kontrolle sowie Möglichkeiten der Generierung von Plots der Geschichte. Im Folgenden werden einige typische Beispiele besprochen.

3.3.2.1 Oz

Das Oz Projekt der Carnegie Mellon University, vgl. [BS89, Bat93, Mat97], befasst sich nicht vornehmlich mit interaktivem digitalem Video, betrachtet jedoch die Vermittlung von glaubhaften, interaktiven Geschichten als seinen Schwerpunkt. Mittels KI sollen glaubhafte Agenten als virtuelle Spieler in einer Virtual Reality Umgebung dem Publikum einen interaktiven und zugleich dramatisch-narrativen Aufenthalt⁷⁰ ermöglichen.

Der Begriff des Dramas einer Geschichte, zusammengesetzt aus glaubhaften Agenten und deren Präsentation, soll an dieser Stelle anhand seiner Bausteine erläutert werden. Glaubhafte Agenten setzen sich im Oz-Projekt aus verschiedenen Faktoren zusammen:

⁶⁷ Vgl. [Joy87].

⁶⁸ Vgl. Kapitel 3.2.1.

⁶⁹ Vgl. Kapitel 3.2.1.

⁷⁰ Die besondere Eigenart dieser Systeme wird nach der Vorstellung von Oz am Ende dieses Kapitels diskutiert.

- Persönlichkeit: Dies beschreibt die einmaligen und spezifischen Eigenheiten, die eine Spielfigur ausmachen, z.B. die Art wie sie spricht oder denkt.
- Emotionen: Spielfiguren drücken ihre eigenen Emotionen aus und reagieren auf die Emotionen anderer Spielfiguren und des Publikums.
- Motivation, pro-aktives Verhalten: Spielfiguren reagieren nicht nur auf Aktivitäten von anderen, sondern haben ihre eigenen inneren Antriebe und Begierden, welchen sie, unabhängig von Interaktionen mit anderen Figuren, nachgehen.
- Veränderung: Spielfiguren haben die Möglichkeit, durch Erfahrung mit der Zeit ihr Verhalten zu ändern.
- Soziale Beziehungen: Eine Spielfigur interagiert mit Anderen (Spielfigur oder Publikum) konsistent, gemäß ihrer gegenseitigen Beziehung. Diese Beziehung kann sich, abhängig von den Interaktionen zwischen den Akteuren, ändern.
- Eine Illusion des Lebendigen⁷¹: Eine gewisse Lebendigkeit des Charakters einer Spielfigur erfordert mehrdimensionale Ziele und Aktionen, ein breites Spektrum an Aktions- und Interaktionsmöglichkeiten und eine zeitlich angepasste Reaktion auf Stimuli der Umwelt.

Nun sollte ein Agent ein gewisses Maß an Intelligenz - und zwar in seinem Verhalten - besitzen. Im Gegensatz zur klassischen KI versucht die verhaltensbasierte KI des Oz Projekts keine perfekte, generalisierte Einzellösung zu generieren, sondern ein breites Verhaltensspektrum, das auf viele Situationen passt. Dabei wird im Wesentlichen auf symbolische Darstellungen des Wissens verzichtet, es werden Zustände abgebildet, die uninterpretiert⁷² bleiben.

Die Geschichte, als weiterer wesentlicher Part des interaktiven Dramas, wird im Oz Projekt mittels eines Plot-Systems dargestellt. Ein Plot stellt hier *einen besonderen Moment*⁷³ einer Geschichte dar. Die Plots umspannen einen diskreten Geschichtsraum⁷⁴. Damit ist die Dramaturgie zwar geprägt durch Branching⁷⁵, doch wird kein Branching zum erstellen von Geschichten benutzt:

Der gewöhnliche Ansatz des Geschichtenerstellens würde eine bevorzugte Erzählsequenz mittels struktureller Vorgaben auf die Geschichtswelt realisieren. Statt dessen gibt der Autor der Geschichte eine *Ausführungsfunktion* vor, welche seine Vorstellungen der Ästhetik der Geschichte beinhaltet.

Die Ästhetik der Ausführungsfunktion wird festgehalten mittels einer Menge von Eigenschaften, die über die Ausführungsfunktion in einer Plot-Permutation überprüft werden

⁷¹ Der Agent wird ohne das Ziel generiert, ihm eine wirkliche Intelligenz, also 'wahres Leben' zu schenken. Im Oz Projekt werden Agenten generiert, um eine Software-Maschine zu besitzen, die das wahre Leben möglichst gut als Illusion darstellt.

⁷² Die Art des Zustands und seine Inhalte werden durch das System nicht interpretiert/evaluiert.

⁷³ In einem Hollywood-Feature-Film gibt es etwa 12-15 besondere Momente.

⁷⁴ Als Geschichtsraum wird dabei eine Menge von Permutation aller möglichen Plots bezeichnet.

⁷⁵ Vgl. siehe Kapitel 3.2.1.

können. Durch die entsprechende Gewichtung der Eigenschaften im Gesamtkontext der Ausführung der Geschichte kann so aus der Menge von Plot-Permutationen zu jedem Zeitpunkt der ideal zur aktuellen Situation passende nächste Plot ausgeführt werden. Die aktuelle Situation erkennt die Ausführungsfunktion im aktuellen Plot unter Berücksichtigung von Publikumsinteraktionen, Spielfiguren und schon besuchten Plots.

Die Ausführung einer Geschichte wird im Oz-Projekt einem *Drama-Manager* übergeben. Dieser Drama-Manager überprüft den Zustand der Geschichtswelt ähnlich einer Spielmaschine: Jede Publikumsinteraktion innerhalb eines Plots wird vom Drama-Manager beobachtet. Eine Sequenz von Ereignissen, welche die Transition zu einem neuen Plot erfordern, werden als Spielzug des Publikums abstrahiert. Der Drama-Manager kann anhand einer Menge von Operationen zur Geschichtswelt-Veränderung seinen eigenen Spielzug ausführen. Dieser Spielzug wird in Ähnlichkeit zu Schach-Spielen generiert. Die möglichen Drama-Manager Spielzüge werden kombiniert mit den darauf folgenden möglichen Publikums-Spielzügen und ergeben die totale Historie von möglichen Spielzügen. Im Bezug auf die bestehende Historie von Plot-Punkten wird dann die Geschichtswelt auf einen Plot hin verändert, der eine maximale Möglichkeit einer sinnvollen totalen Historie gewährt.

Der Ansatz von Oz trennt Charaktere (Agenten) von der Erzählweise der Geschichte. Dies führt zu einer Echtzeit-Generierung von Verhalten für glaubhafte Agenten, die dem Konzept von digitalem Video fremd ist und als generisches Konzept zur Narration von Geschichten zu werten ist. Aufgrund des Einsatzes des Drama-Managers erfolgt die Kontrolle der Geschichte global - die Granularität der Kontrolle erfolgt dabei über das gesamte Spektrum der Möglichkeiten von feiner Kontrolle der Möglichkeiten eines Agenten bis zur groben Kontrolle der Plots. Das Oz System kann in die Interaktive Drama Kategorie eingeordnet werden.

3.3.2.2 Virtual Theatre and IMPROV

Das Virtual Theatre Projekt der Universität Cambridge, vgl. [Spr98], ist der Versuch, die Konvergenz von verschiedenen Medien wie TV, Kino, Multi Media, über einen kollaborativen, dramatischen Ansatz für nichtlineare Geschichten zu erreichen. Der Ansatz von Virtual Theatre besteht darin, das Publikum nicht als Autoren, sondern als Spielfigur in die Handlung zu integrieren - in Kollaboration mit anderen Spielfiguren, welche ebenfalls aus dem Publikum gestellt werden. Damit unterscheidet sich Virtual Theatre im Ansatz von Systemen wie z.B. Oz, die auf der Dramatisierung einer einzelnen Geschichte für ein einzelnes Individuum basieren.

Das entscheidende Prinzip von Virtual Theatre ist dabei die Bereitstellung einer Umgebung, die ein gewisses dramatisches Potential enthält und die Situation der Spielfiguren vorantreibt. Dieses Potential kann zum Beispiel aus Intrigen, Gegnern, Beeinflussung von Außen oder auch einer Hintergrundgeschichte bestehen. In Virtual Theatre fehlt die Möglichkeit, tatsächlich die Narration einer Geschichte gezielt zu beeinflussen, da die Narration komplett von der Interaktion der Spielfiguren untereinander abhängt.

Ähnlich gestaltet sich das System IMPROV, vgl. [PG98]. IMPROV erlaubt eine sehr fei-

ne, mehrstufige Verhaltensabstimmung von Agenten mittels Scriptsprachen, die speziell für eine entsprechende Verhaltensbeschreibung entwickelt wurden. Tatsächlich ist der dramaturgische Ansatz sehr ähnlich zu Virtual Theatre, da die Narration einer Geschichte ausschließlich mit dem Verhalten der Spielfiguren und der Interaktion des Publikums mit den Spielfiguren vorangetrieben wird. Eine gezielte Geschichte ist somit eher schwierig zu erzählen.

Beide Ansätze sind generative Systeme, welche der Interactive Fiction zuzuordnen sind. Eine grobe Kontrollgranularität ist nicht gegeben, die Kontrolle der Narration ist ausschließlich Lokal angesiedelt.

3.3.2.3 Codename: Extinct

Codename: Extinkt ist ein Spiel, das durch eine besondere Nutzung von interaktiven Geschichten auffällt. Der Anspruch von Codename: Extinct ist die Qualität eines Feature-Films bezüglich des *Looks* und des *Feels*. Die Problematik des *Feels* soll hier näher erläutert werden.

Heintze, vgl. [vHBB00], beschreibt das Entstehen von Spannung im klassischen Film als Fähigkeit des Autors, gewisse Lücken zwischen Publikumserwartung und Auflösung der Geschichte aufzubauen. So werden große Teile einer Handlung durch den Autor offenbart, die entscheidende Wendung jedoch versteckt und unerwartet präsentiert. Um eine Spannung aufzubauen ist es also nicht ratsam, dem Publikum die Macht eines Regisseurs/Autors und damit das gesamte Wissen über eine Geschichte in die Hand zu legen. Dem Benutzer bleibt die Möglichkeit als mitwirkender Akteur in die Handlung einzugreifen. Durch seine Interaktionen kann eine spezielle Erwartungshaltung abgeleitet werden, welche die Spielmaschine zum Entwerfen von unerwarteten Ereignissen nutzen kann. Diese Ereignisse müssen allerdings innerhalb der Constraints des Autors und dramatisch korrekt präsentiert werden. Somit wird die Geschichte dramatisch, während das Konzept selbst eine Nichtlinearität, d.h. eine variable Geschichte, hervorbringt.

Obige Theorie verlangt eine besondere Herangehensweise an eine Geschichte: Statt die interaktive Geschichte mit Hilfe des Schemas Branching bzw. String of Pearls⁷⁶ zu erzählen, welche eine fest vernetzte Narrationsstruktur durch eine Anzahl sogenannter Nodes besitzen, nutzt das Spiel die Methode der *Dynamischen Nodes*. Hierbei wird kein festes Netzwerk von Nodes benutzt, an welchem die Narration sich weitgehend orientiert. Interagiert das Publikum mit der Geschichte, so wird diese automatisch, z.B. durch Veränderung der Rolle des Protagonisten, angepasst. Zu diesem Zweck wird anhand der Interaktion des Benutzers dessen Erwartungshaltung impliziert und ein entsprechender dramaturgischer Effekt geplant. Mit dem dramaturgischen Effekt wird daraufhin der nächste Node geplant.

Der Dramaturgische Aufbau von Codename: Extinkt ähnelt dem Drama-Manager des Oz-Projektes. Codename: Extinkt rückt allerdings den Suspense des Spielers in den Fokus der dramaturgischen Generierung der Narration, während beim Oz-Projekt die sinnvolle totale

⁷⁶ Vgl. Kapitel 3.2.1.

Historie von Nodes (entsprechen Plots im Oz-Projekt) im Vordergrund des Interesses steht.

3.3.3 Zusammenfassung

Nach der Betrachtung von sowohl VR-basierten, generischen Ansätzen und videobasierten, selektiven Ansätzen, lässt sich folgern, dass die reine Präsentation von Video einem gehobenen Anspruch des Storytelling (generische Geschichten, variables Informationsmanagement, generative Kontrolle der Geschichte, interaktiver Zugang zur Kontrolle von Storyelementen, Kameraführung) nicht entspricht. Trotzdem besitzt der Einsatz von Video Vorteile, die es rechtfertigen, mit diesem Medium zu arbeiten: Video kann sehr effizient zur Einschränkung eines Informationsraumes durch einen Autoren genutzt werden. Dies ist eine wertvolle Hilfe bei der Beschränkung der durch das Publikum geforderten Narrationsmöglichkeiten mit immersiven Mitteln. So kann Video als Kontextgebendes Medium bei medienzentrierten Interaktionen vom Publikum sehr gut verstanden und damit genutzt werden.

Aus den Vorteilen der VR-basierten Ansätze und den videobasierten Ansätzen der nichtlinearen Narration lässt sich der Anspruch eines kombinierten selektiv/generischen, interaktiven Narrationssystems ableiten, welches auf Video basiert, jedoch VR-Elemente zur Erweiterung der Narrationsmöglichkeiten benutzt. Dieser Anspruch führt im nächsten Kapitel zur Formulierung der entsprechenden Anforderungen an nichtlineares Video.

3.4 Anforderungskatalog für nichtlineares Video

An dieser Stelle soll die Zielsetzung dieser Arbeit und damit die Anforderungen an ein Narrationssystem für nichtlineares Video, die der Erfüllung der Arbeitshypothesen⁷⁷ dienen, explizit präzisiert werden.

Die verschiedenen Aspekte und Zielsetzungen der nichtlinearen Erzählstruktur werden im Folgenden benannt:

- Autoren-Zugänglichkeit: Aufgrund der fehlenden allgemeinen Gültigkeit eines separaten Geschichtenmodells muss die Erzählstruktur von Autoren anpassbar sein.
- Freiheitsgrad der Interaktion durch den Benutzer: Die Handlungen des Spielers sollten tatsächlich einen signifikanten Einfluss auf die erzählte Geschichte nehmen. Dies geht über das potentielle Ein- oder Ausfügen von Geschichtesteilen hinaus, hin zur Bildung von Geschichtsvarianten. Dies sollte unbedingt so geschehen, dass der Benutzer nicht aus der Geschichtswelt herausgerissen wird, die Verzweigungen sollten also für den Benutzer transparent sein.
- Spannung (Suspense): Das Ziel einer Erzählstruktur ist schon seit Aristoteles der Aufbau von Spannung beim Zuschauer. Dies darf bei der nichtlinearen Geschichte keinesfalls dem Zufall überlassen werden, sondern muss dynamisch auf die Interaktionen

⁷⁷ Vgl. Kapitel 1.1.

des Benutzers angepasst werden, um sowohl der Interaktion als auch den Suspense-Vorgaben zu genügen. Neben dieser Anforderung muss die Suspense-Struktur unbedingt für einen Autoren nachvollziehbar sein, da dieser anhand der Struktur eine entsprechende Geschichte aufbauen sollte.

- Abgeschlossenheit und Zeitmanagement: Ein vorzeitig erzwungenes Ende einer Geschichte oder eine ziellose Navigation (z.B. durch ein Hypertextdokument) haben einen fatalen Effekt auf die Zufriedenheit des Benutzers, vgl. [BWK92]. Sollte eine Erzählzeit vorgegeben sein, so muss gewährleistet sein, dass die Geschichte *auf den Punkt* zu einem dramaturgisch sinnvollen Abschluss gelangen.
- Identifikation und Immersion: Um dem Benutzer die Gelegenheit zu geben, sich mit seiner Rolle innerhalb einer interaktiven Erzählung zu indentifizieren, ist ein gewisses Maß an gedanklicher Immersion in die Geschichte notwendig. Diese Ziele sind z.Z. nicht direkt beschreibbar, da die Grundlagen der gedanklichen Immersion nicht erkannt sind und in der Spannung der Geschichte bzw. in dem allgemeinen Interesse des Benutzers an der Thematik der Geschichte gesucht werden.
- Adaption an einzelne Benutzer: Eine Anpassbarkeit an den Benutzer ist von besonderer Bedeutung für eine Erzählstruktur im allgemeinen - so sind z.B. für verschiedene Altersgruppen unterschiedliche Inhalte in verschiedenen Häufigkeiten präsentierbar (z.B. wird eine Geschichte für Kinder durch extreme Wiederholung der Inhalte geprägt). Neben diesen psycho-sozialen Kriterien sind natürlich auch moralische Kriterien wie Gewalt oder Sexualität im Hinblick auf eine Benutzeradaption interessant. Bei der nichtlinearen Geschichte kommen augenscheinlich noch Faktoren wie Interaktionsfreiheit, Schwierigkeitsgrad der Spielbarkeit und Handlungsfülle hinzu.

Anhand dieser Ziele erfolgt eine kritische Diskussion der bestehenden Ansätze⁷⁸.

- Charakter-basierte Ansätze: Diese können nicht die Spannung und Handlungskohärenz erreichen, die in den Zielen⁷⁹ gefordert sind.
- Emergent Narrative: Da sehr eng verwandt mit charakterbasierten Ansätzen fehlt auch dem Prinzip des Emergent Narrative die notwendige Struktur, um spannende und darstellenswerte Handlungen von uninteressanten Details zu unterscheiden. Dies ist aber ein sehr dringendes Ziel des Geschichtenerzählens: Gerade das Entfernen von sogenanntem *Alltagsrauschen* aus einer Geschichte macht diese interessant und einprägsam.
- Regelbasierte Ansätze: Der Versuch, die Kreativität eines Menschen in eine Menge von Regeln zu pressen und daraus Geschichten erzeugen zu lassen, ist nach Bringsjord und Ferrucci Zwangsläufig zum Scheitern verurteilt, vgl. [BF99]. Crawford, vgl. [Cra99], schließt ebenfalls aus, dass es eine algorithmische Lösung zur synthetischen,

⁷⁸ Vgl. Kapitel 3.2.

⁷⁹ Vgl. Kapitel 3.2.

vollautomatischen Generierung von Geschichten geben kann - er fordert die Anerkennung des menschlichen Autors und die Unterstützung desselben mittels Werkzeugen, die seine kreative, geschichtenerzeugende Aufgabe erleichtern. Hier erfolgt der Schluss, dass durchaus ein Regelwerk zur Ablaufsteuerung einer Geschichte sinnvoll einzusetzen ist - soweit die einzelnen Teile der Geschichte bekannt und bewertet sind, im Sinne einer funktionalen Bewertung - d.h.: welche Funktion erfüllt eine Geschichte bei der Partizipation der Geschichte durch den Benutzer?

- Weltbasierte Ansätze: Die Problematik von weltbasierten Ansätzen liegt in der Komplexität der Simulation der entsprechenden Storywelt. Um eine Storywelt zu simulieren müssen unnötig viele Details der Story beschrieben werden. Dies behindert den Autoren in der Konzentration auf die wesentlichen dramaturgischen Elemente seiner Geschichte und überlassen den Suspense dem Zufall.

Zusammenfassend lässt sich urteilen, dass eine Definition der Story-Handlung bis ins letzte Detail nicht sinnvoll ist, da dies einerseits das Authoring von Geschichten erschwert oder verhindert und andererseits das Hauptaugenmerk der Ansätze auf Simulation und nicht auf Narration richtet. Ein regelbasierter Ansatz der Steuerung des übergeordneten Handlungsverlaufs einer (von Menschen generierten und funktional bewerteten) Geschichte verspricht jedoch die Ziele der nichtlinearen Narration zu erfüllen. Aufgrund der Unzulänglichkeiten der bekannten Modelle bezüglich der aufgestellten Ziele der nichtlinearen Narration erfolgt eine Anpassung des regelbasierten Konzeptes mittels der Morphologie von Propp⁸⁰. Diese entspricht in der Hierarchie von Aylett, vgl. [Ayl99], der mit *übergeordnete Handlung* bezeichneten obersten Ebene. Handlungssequenzen auf Charakterebene werden aus einzelnen Szenen und deren Inhalten gewonnen - diese sind am zu präsentierenden Content (etwa Video-Clips) und der zugehörigen (konversationalen) Interaktion orientiert - und werden als zweite Ebene bezeichnet.

Für ein zugrundeliegendes Narrationssystem wurde in diesem Kapitel gezeigt, dass zumindest ein Branching-basierter Aufbau des Videos unterstützt werden sollte. Die Zusammenstellung von Clips zu nichtlinearen Präsentationen sollte wenigstens halbautomatisch (d.h. durch Assistenz des Autors/Anwendungsprogrammierers), im günstigsten Falle automatisch (nach Autoren- und Publikumspräferenzen) gestaltet sein. Um die ansatzbedingten Einschränkungen eines selektiven Ansatzes für nichtlineares Video (keine Neu-Generierung von Geschichten) zu entschärfen, wird ein generativer Zusatz zum Video favorisiert, der dem Publikum mittels Animation eine zusätzliche, frei gestaltbare Möglichkeit der emotionalen und inhaltlichen Präsentation sowie konversationale, personalisierte Interaktion bietet.

Die Präsentation des Videos und der angegliederten, animierten personalisierten Interaktionsmöglichkeit sollte medienzentriert vonstatten gehen, im Falle von Video getrieben durch die Inhalte des Videos und durch den Aspekt der zeitlichen Darstellung von Inhalten. Dies erfordert einmal eine Interaktion mit den visuell und akustisch dargestellten Objekten/Inhalten des Videos, die eine direkt vom Autor vorgegebene narrative Änderung der Informationsdarbietung ermöglichen. Auf der anderen Seite soll die Metapher der kon-

⁸⁰ Vgl. Kapitel 3.2.2.2.

versationalen Interaktion eingesetzt werden, um einerseits einen festen Ansprechpartner bieten zu können und andererseits eine nicht zeitgebundene Delegation zu gewährleisten, die es dem Publikum ermöglicht, auf Wunsch narrative Änderungen in der Präsentation des nichtlinearen Videos vorzunehmen. Das Video gibt dabei den Kontext der Konversation im Wesentlichen vor, wird daher zur Einschränkung der Anforderungen des Publikum an die Informationsdarstellung genutzt. Der konversationale Ansprechpartner sollte sowohl als Benutzerschnittstelle als auch als Teil der Videopräsentation arbeiten. Dies bedingt eine feine Mimik und Gestik, welche möglichst eng an das menschliche Vorbild angepasst ist.

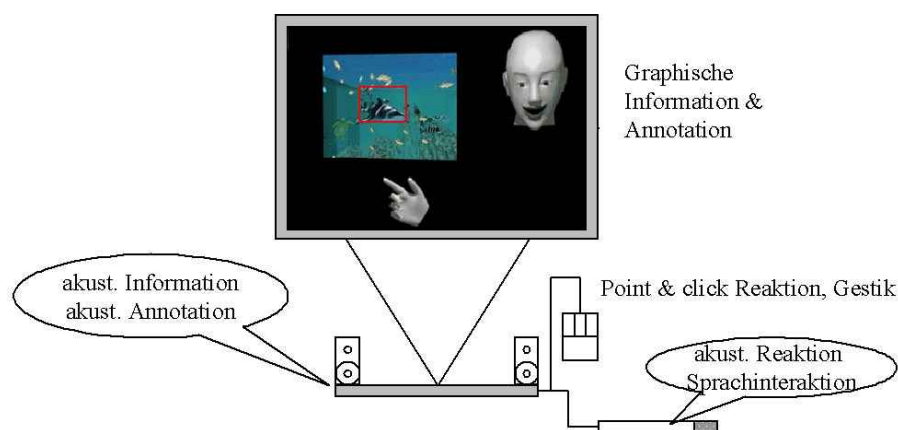


Fig. 3.9: Synchroner und Asynchroner Interaktionswerkzeuge

Abbildung 3.9 zeigt, wie akustische und graphische Informationen und Annotationen in Kombination genutzt werden. Das Publikum reagiert mit entsprechenden graphisch-interaktiven bzw. akustisch-interaktiven Mitteln.

Die gesamte Präsentation des Videos sowie die zugehörigen Interaktionsmöglichkeiten zur Beeinflussung der Narration durch Autor und Publikum müssen einem entsprechend hohen Designanspruch genügen, um die Immersion des Publikums in die Geschichte nicht störend zu beeinflussen.

Aus der Kombination der Autoren- und Präsentationsanforderungen entsteht die Vision eines kombiniert selektiven/generativen Storytelling-Ansatzes mit der Möglichkeit zur kombiniert direkt-manipulativen und konversationalen narrativen Beeinflussung der Geschichte, die durch nichtlineares Video erzählt wird, siehe Abbildung 3.10. Als selektives Grundmedium wird Video angestrebt, erweitert um die generative Komponente des Avatarverhaltens und der sonstigen Zusatzinformationen. Die Interaktionsangebote werden direkt vom Video als kontextvorgebendes Medium angetrieben, das Publikum kann über Sprache und Zeigegestik (Maus)⁸¹ interagieren.

Die oben notierten Anforderungen ergeben sich, um die Erfüllung der in Kapitel 1.1 gestellten Arbeitshypothesen zu gewährleisten. In den folgenden Kapiteln wird ein solches nichtlineares Videosystem, vgl. [Bra01a], aufgespannt, die Techniken zu dessen Implemen-

⁸¹ Diese Konzepte werden in Kapitel 4 eingeführt und diskutiert.

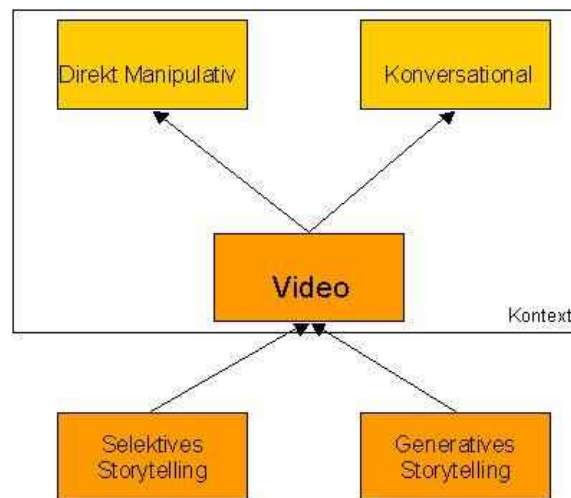


Fig. 3.10: Modell eines generativ/selektiven Storytelling-Systems, basierend auf nichtlinearem Video, mit direkt manipulativem und konversationalem Interaktionszugang

tierung erläutert, eine Verifikation des Ansatzes durchgeführt sowie die Arbeitshypothesen dadurch als zutreffend gezeigt.

3.5 Zusammenfassung

Das Kapitel stellt die Grundlagen und Ideen von Storytelling vor. Anhand von notierten Eigenschaften und Kriterien des Nonlinear Storytelling werden generelle Ansätze diskutiert sowie aktuelle Forschungsansätze und Konzepte sowohl in Bezug zu Digitalem Video als auch Virtual Reality bewertet. Aus den daraus gewonnenen Einsichten werden Anforderungen für ein Video-basiertes, interaktives System zur nichtlinearen Narration abgeleitet. Diese Idee eines kombinierten selektiv- generativen Narrationssystem basierend auf Video, integriert in eine konversationale- und direkt manipulative Benutzerumgebung findet sich in keinem der in Kapitel 3.3.1 und Kapitel 3.3.2 beschriebenen Ansätze wieder. Aus diesem Grund erscheint der Aufbau eines solchen Systems, wie in Kapitel 3.4 beschrieben, geradezu zwingend. Der Ansatz dieser Arbeit sollte die verschiedenen Interaktions- und Storytellingmetaphern zusammen führen und damit sowohl für Publikum als auch Autoren eine neue Dimension der nichtlinearen Geschichtsnarration basierend auf Digitalem Video erschließen. Durch die angestrebte generische Geschichtsautomatisierung in Kombination mit kontextgetriebener Konversation und direkter Manipulation ist der Autor einer nicht-linearen Geschichte frei in der Definition seiner Ästhetik als auch seiner Narrationsweise. Dem Publikum scheint die Benutzerschnittstelle zunehmend transparent und zentriert auf den Umgang mit dem nichtlinearen Medium, vgl. [Bra01a].

4. INTERAKTION FÜR KONTINUIERLICHE MEDIEN

Möglichkeiten der Interaktion haben einen starken Einfluss auf das Erleben von Geschichten: Neben dem Einfluss auf den Erzählfluss selbst kann die technisch-gestalterische Durchführung der Interaktion die Immersion in eine Geschichte stark beeinflussen. Die Interaktionsmöglichkeiten können wie folgt unterschieden werden:

- Direkt Manipulativ: Eine Interaktion bezüglich eines Objektes der Geschichte wird angeboten. Diese Interaktion wird mittels eines Werkzeugs direkt am Objekt durchgeführt.
- Delegation & Assistenz: Eine Aufgabe wird nicht durch den Benutzer am Objekt selbst mittels entsprechender Werkzeuge durchgeführt, sondern durch einen Mittler, der die entsprechende Aufgabe durch den Benutzer erhält (und eventuell auch eine Erfolgs- bzw. Misserfolgsmeldung an den Benutzer gibt).

Nach Bahrtdt, vgl. [Bah90], wird in der Soziologie als Interaktion ein wechselseitiges soziales Handeln von zwei oder mehr Personen bezeichnet. Bahrtdt fordert, dass sich jeder Partner in seinem Handeln daran orientiert, dass der andere sich in seinem Handeln auf das vergangene, gegenwärtige oder zukünftige Handeln des ersteren bezieht. Nach Bahrtdt gehört zur Interaktion ferner ein Konsens über ein gemeinsames unmittelbares Handlungsziel. Interaktion bedarf des Informiertseins über die Intentionen des jeweils anderen, vergl. auch Bibel [Bib03, 95pp].

Der direkt-manipulative Ansatz der Interaktion ist dem Benutzer durch Präsentation der entsprechenden Werkzeuge einfach zu vermitteln. *Einfach* bedeutet in diesem Zusammenhang, dass der Benutzer durch die Werkzeuge in seiner Interaktionsfähigkeit begrenzt ist. Alle Interaktionsvorhaben des Benutzers, die sich nicht mittels der Werkzeuge durchführen lassen, sind für den Benutzer vorweg zu erkennen. Diesem Ansatz fehlt allerdings die u.a. von Bahrtdt geforderte soziale Komponente.

Im Gegensatz dazu ist beim Einsatz von Assistenz und Delegation für den Benutzer oft schwer zu erkennen, was er delegieren kann und vor allem wie er es delegieren kann. Zur Delegation werden oftmals natürliche, dem Benutzer bekannte Methoden, wie zum Beispiel natursprachliche Formulierungen, benutzt. Die Problematik diesbezüglich ist, dass dem Benutzer suggeriert wird, jede natursprachlich zu formulierende Delegation auch wirklich formulieren zu dürfen - und damit einem entsprechenden System zur Aufgabe zu stellen. An dieser Stelle ist es wichtig, dass der Autor durch spezielle Vorkehrungen, zum Beispiel durch die Vorgabe eines gewissen Kontexts und damit einer Reduzierung der natursprachli-

chen Möglichkeiten, die Delegationswünsche des Benutzers in für den Autor beherrschbare Bahnen zu bringen. Positiv wirkt sich mit dieser Interaktionsform die Nutzung von Konversation aus, da sie u.a. die von Bahrddt geforderten Elemente der Interaktion mit einbezieht. In diesem Kapitel werden drei Begrifflichkeiten verwendet, welche von einer grundlegenden Bedeutung für das Verständnis dieses Kapitels sind - aus diesem Grund werden sie vorab definiert:

Medium Das Medium ist ein Vermittlungssystem für Informationen aller Art, ihre Funktion ist der Transport von Inhalten. Steinmetz [Steinmetz:B95] unterscheidet folgende (technischen) Arten von Medien:

- Perzeptionsmedien (Sinnesorgan): visuell (Text, Einzelbild, Bewegtbild) oder auditiv (Musik, Geräusch, Sprache).
- Repräsentationsmedien (rechnerinterne Codierung): z. B. Text im ASCII-, Grafik im JPEG-Format.
- Präsentationsmedien (Ein- und Ausgabe): Eingabemedium (Tastatur, Kamera, Mikrofon) und Ausgabemedium (Bildschirm, Papier, Lautsprecher).
- Speichermedien (Datenträger): z. B. Diskette, Festplatte, CD-ROM, DVD, Papier.

In dieser Arbeit wird der Begriff im Sinne des Perzeptionsmediums verstanden. Werden andere Arten von Medialität (z.B. der Fernseher als Präsentationsmedium) angesprochen, so ergibt sich dies aus dem Kontext der Beschreibung.

Modalität Eine Modalität ist ein Reiz¹, der mit einem Sinnesorgan des Menschen erfasst werden kann.

Kodalität bezeichnet das Symbolsystem, das die Symbole definiert, mittels derer eine Information beschrieben ist. Ein Beispiel soll diese Begrifflichkeit weiter verdeutlichen: Die Aussage: *Ich weiß das nicht.* kann als Text (wie hier geschehen), als Mimik (Gesicht mit hochgezogenen Augenbrauen und hängenden Mundwinkeln), als Gestik (erhobene Arme in L-Form mit nach oben geneigten, geöffneten Handflächen) oder als Piktogramm (Fragezeichen-Symbol) kodiert sein.

Wird der Begriff *Multi* im Zusammenhang mit den oben definierten Begriffen genutzt, so bedeutet dies, dass mehrere Ausprägungen der oben genannten Begrifflichkeiten in Kombination gezielt angewendet werden - so bedeutet Multimedia, dass mehrere Medien - oftmals wird ein kontinuierliches und ein diskretes Medium gefordert - kombiniert zur Darstellung von Informationen verwendet werden.

Grundlegend für dieses Kapitel werden die Begriffe Hypermedia und Hyperlink sowie Annotation verwendet.

¹ Eine erweiterte Definition der verschiedenen Reizmöglichkeiten findet sich in Kapitel 4.1.2.

Hypermedia Als Begriff eingeführt von Ted Nelson, vgl. [Nel92], bezeichnet Hypermedia ursprünglich Hypertext-Systeme, welche verschiedene Medien miteinander verknüpfen. Der Begriff Hypertext selbst ist von Nelson 1965 eingeführt worden. Hypertext bezeichnet Systeme, welche Informationen in einem Netzwerk von Nodes mittels Verknüpfungen miteinander in Beziehung setzen. Entgegen der traditionellen linearen Strukturen enthalten Hypertexte also Verzweigungen (sog. Links oder Hyperlinks), die nur auf Aufforderung sichtbar werden.

Hyperlink Eine Beziehung von einer Datenquelle zu einer Datensenke. Der Hyperlink ist eine spezielle Form der Annotation eines Mediums.

Annotation Der Begriff bezeichnet eine (sichtbare) Markierung eines Mediums oder eines speziellen Teils eines Mediums.

In dieser Arbeit wird für kontinuierliche Medien im Allgemeinen und am Beispiel von Digitalem Video im speziellen eine State of the Art Analyse zu den folgenden Thematiken durchgeführt:

- Akustische Informationsdarstellung und Interaktion
- Graphische Informationsdarstellung und Interaktion
- Konversationale Interaktion

Das Ziel dieses Abschnittes ist die Vorstellung von technischen Ursprüngen und Anknüpfungspunkten für die Forschungen bezüglich nichtlinearen Erzählstrukturen mit interaktivem Video, welche in dieser Arbeit dargestellt sind. Die drei Themen werden unter Anderem im Bezug auf Präsentation und Authoring der entsprechenden Daten diskutiert. Tatsächlich sind Konversationen prinzipiell als kontinuierlich, d.h. auf einer Zeitachse stattfindend, anzusehen. Aus diesem Grund werden sie in diesem Kapitel noch vor den akustik- und grafikbezogenen Interaktionen behandelt, die sowohl einen kontinuierlichen (und konversationalen) als auch einen diskreten Charakter haben können.

4.1 Konversationale Interaktion

Konversationale Benutzungsschnittstellen simulieren eine menschähnliche Art der Kommunikation zwischen Benutzer und System. Es drängt sich an dieser Stelle die Frage auf, was ‚menschähnlich‘ bedeutet. Der Begriff ‚menschähnlich‘ ist sehr breit gefasst, bezüglich Kommunikation bedeutet er die Simulation eines natürlichen Gesprächs zwischen zwei Menschen, wie es in Abbildung 4.1 gezeigt wird. Der Austausch zwischen den beiden Parteien muss dabei nicht über gesprochene Sätze erfolgen, sondern bezieht sich auf jedwede Medialität.

Eine Konversation wird in der vorliegenden Arbeit durch folgende Faktoren bestimmt - die Begrifflichkeit des Satzes wird dabei stellvertretend für sonstige Medien ausgelegt:

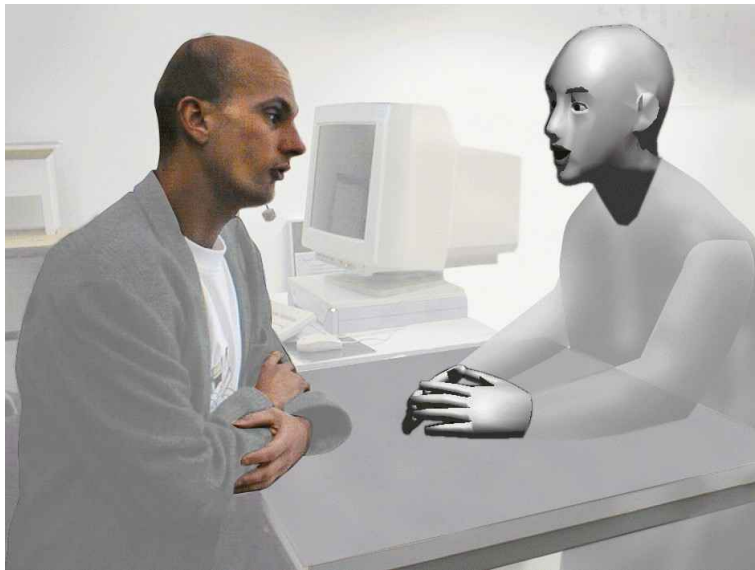


Fig. 4.1: Menschähnliche Art der Kommunikation

- Eine Konversation findet zwischen zumindest zwei Konversationspartnern statt. (Ein Konversationspartner stellt eine - durch den Menschen wahrnehmbare - Einheit dar.)
- Eine Konversation findet im Rahmen eines Informations-Kontextes statt. Dieser Kontext wird oftmals nicht explizit benannt, allerdings beziehen die gesprochenen Sätze den Kontext immer mit ein.
- Eine Konversation bedeutet ein wechselseitiges Sprechen, Zuhören und Verstehen.
- Eine Konversation hat eine inhaltliche, aber auch eine soziale und emotionale Komponente.
- Eine Konversation hat eine zeitliche Ausdehnung, daher besitzt sie einen Anfang, eine Dauer und ein Ende. Konversation ist ein kontinuierlicher Vorgang, die sich innerhalb eines zeitlichen Intervalls abspielt.

Die Interaktion², die während einer Konversation zwischen System und Benutzer stattfindet, wird im Feld der Sprachtheorie durch Speech Acts, vgl. [Gri75], beschrieben. Diese Speech Acts beschreiben semantische Konversationskomponenten. Speech Acts werden problembezogen und damit applikationsspezifisch definiert. Ein Speech Act muss nach Austin philosophisch unter folgenden Aspekten betrachtet werden, vgl. [Bac00]:

² Der Begriff der Konversationalen Interaktion wird in dieser Arbeit unabhängig von der medialen Ausprägung der Interaktion gesehen. Aus diesem Grund wird in diesem Kapitel nicht auf das Verstehen und die Generierung von Sprache eingegangen, sondern auf die symbolisch-abstrakte Komponente - das conversational gambit - der Konversation, siehe Kapitel 5.2.2. Diese Komponenten werden zwar im User Interface ebenfalls auf Modalitäten abgebildet, die Modellierung erfolgt jedoch explizit ohne Berücksichtigung der Modalitäten.

Lokutionärer Akt Der Akt des Aussprechens eines Inhaltes.

Illokutionärer Akt Die kommunikative Absicht des lokutionären Akts.

Perlokutionärer Akt Der Wirkungsaspekt des Sprechaktes.

Die Begrifflichkeiten werden am Beispiel einer Sprechakt-Analyse eines Satzes erläutert:
(Kontext: Mann spricht zu Frau, Frau 5 Minuten zu spät): *Wie viel Uhr ist es?*

Lokutionär Aussprechen der Aussage.

Illokutionär Vorwurf

Perlokutionär Bewirken, dass die Frau sich beeilt.

Übertragen auf die Mensch Maschine Schnittstelle bedeutet dies, dass eine mediale Aussage von der Maschine nicht nur lokutionär, sondern auch illokutionär gedeutet werden muss, um die perlokutionäre Absicht berücksichtigen und ausführen zu können.

4.1.1 KI und Sprachtheorie, Bottom Up Ansatz

Konversation wird im Bereich der Künstlichen Intelligenz (KI) oft mit der Verarbeitung von Natursprache gleichgesetzt. Sprachverarbeitung gliedert sich in die Bereiche Spracherkennen und Sprachverstehen, vgl. [Har85]. Sprachverstehende Systeme sind im allgemeinen hierarchisch aufgebaut und bestehen aus den Ebenen Phonetik, Syntaktik, Semantik und Pragmatik. Im Allgemeinen steht hinter einem sprachverstehenden System die Beschreibung einer Sprache mittels einer Grammatik, die diese Sprache umfassend beschreibt. Da eine grammatische Beschreibung endlich ist, versteht sich von selbst, dass kein sprachverstehendes System wirklich jedes Wort bzw. jeden Satz verstehen kann. Somit ist der Benutzer immer auf die Sprachmenge beschränkt, der durch die Grammatik des Systems vorgegeben wird.

Der Interpretationsspielraum bezüglich des Kontextes bei natursprachlicher Kommunikation ist recht groß. Aus diesem Grund wird in der KI durch die Definition von dem System bekannten Speech Acts der Interpretationsraum verkleinert. Dies erfolgt in Einklang mit der Konversationsimplikaturtheorie des Linguisten H. Paul Grice, vgl. [Gri75], welcher vier grundlegende Konversationsaxiome (Maxime) formulierte:

Quantity Inhalte sollten so informativ wie nötig und so simpel wie möglich gestaltet sein.

Quality Inhalte sollten dem entsprechen, was der Konversationsteilnehmer für wahr hält.

Relevance Die Konversationsbeiträge sollten ein Thema auf den Punkt bringen.

Manner Der Konversationsbeitrag sollte klar verständlich sein.

Das damit beschriebene Prinzip *Be Cooperative* fordert dazu auf, Beiträge zur Konversation so zu gestalten, wie es die Ausrichtung und Zweckbestimmung des Gespräches erfordert.

Offensichtlich ist der Begriff des Speech Acts sehr stark von der KI geprägt und vor allem auf Sprachverstehen und Sprachgenerierung ausgelegt. Diese Auslegung ignoriert jedoch den Umstand, dass Konversation nur zum Teil aus den kommunizierten Inhalten (Inhalt als Content, d.h. das Thema der Konversation) der Konversation besteht. Der zweite Faktor der Konversation bezieht sich auf das konversationale Verhalten im Allgemeinen. So beschreibt Wahlster, vgl. [Wah91], bezüglich Spracherkennungssystemen eine Erhöhung der Robustheit durch das Einbeziehen von Gestiken des Benutzers.

Crangle, Carlson und Tuttle, vgl. [CFC⁺97], erkennen diesen Faktor und versuchen, den Sprach-basierten Ansatz um fünf Prinzipien für natürlichsprachliche Benutzerschnittstellen zu erweitern:

1. Recognize the contributions of all interface elements: Das Prinzip fordert die Einbindung weiterer Modalitäten (neben Sprache) in die Konversation.
2. Establish common ground for prompts: Allgemeines Wissen über den Kontext der Konversation sollte bei der Generierung von (Sprach-) Ausgaben berücksichtigt werden.
3. Syntactically and lexically encode the form of the desired reply in the verbal prompt: Die Ausgabe sollte von der Wort- und Themenwahl des Konversationspartners beeinflusst werden.
4. Establish a canonical form of interaction: Verbindliche Richtlinien (z.B. der Wortwahl) sollten für jeden Konversationspartner gelten.
5. Allow the user to exercise familiar conversational gambits: Der Benutzer sollte die Möglichkeit haben, typische Mensch-Mensch-Kommunikation auch dem System gegenüber anzuwenden - zum Beispiel das Unterbrechen des Systems zu einem beliebigen Zeitpunkt.

Diese fünf Prinzipien sind tatsächlich immer noch sehr stark an die Sprachverarbeitung angelehnt. Das eigentliche zentrale Prinzip des *conversational gambit*, d.h. des Anwendens von menschlichen Verhaltensweisen in der Konversation - wird zwar erwähnt, jedoch in seiner Bedeutung relativiert. Weitere Arbeiten auf diesem Gebiet, zum Beispiel Wahlster, vgl. [Wah02], greifen ebenfalls die Kombination von Sprache und Mimik als zielführend bei der Kommunikation mit Computersystemen auf.

4.1.2 Konversation, Top Down Ansatz

Der Unterschied zwischen einem klassischen Dialog des Benutzers mit einem Rechner und einer Konversation liegt im Kontext begründet. Der Kontext eines Rechner-Dialoges ist stark eingeschränkt, während der Kontext einer Konversation eher breit angelegt ist, vgl.

Bibel [Bib03, p.98]. Wird die Berücksichtigung eines Kontextes zur Kommunikation mit dem Benutzer gefordert, so kann dies mit dem Begriff der Konversation beschrieben werden. Billinghurst, vgl. [CBB⁺99], fordert die Einbringung von Ansätzen aus den Bereichen Multimodale Interfaces, Konversationale Modelle und körperliche, konversationale Interfaces zur effektiven Konzeption und Darstellung von Konversation, siehe Abbildung 4.2³. Dies folgert Billinghurst direkt aus Beobachtungen menschlicher Kommunikation. Diese ist sowohl multimodal, folgt gewissen Konversationsprinzipien und besteht zu einem großen Teil aus Körpersprache (Mimik, Gestik).

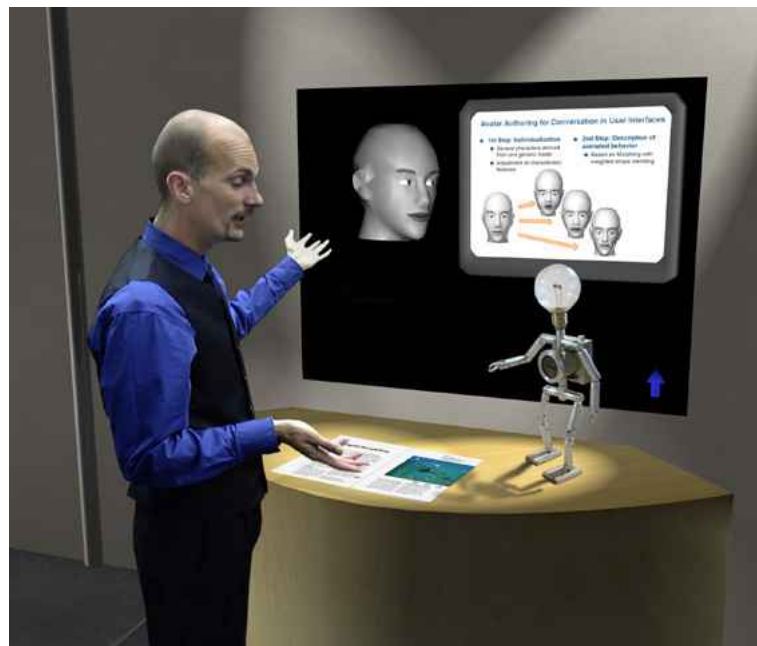


Fig. 4.2: Konversationale Interaktion in einer Mixed Reality Umgebung

Eine Konversation besteht aus zumindest zwei inhaltlichen Komponenten, vgl. [CBB⁺99]:

- Propositionale (aussagebezogene) Komponenten, die sich auf die Inhalte der Konversation beziehen.
- Diskurs-bezogene Komponenten, die sich auf das Management der Konversation beziehen, d.h. auf die Führung eines Gespräches, zum Beispiel ein Turntaking-Signal.

Konversationale Benutzerschnittstellen bedienen sich im Allgemeinen eines Avatars. Konversationale Benutzerschnittstellen in dieser Arbeit unterscheiden sich von sonstigen konversationalen Ansätzen durch den expliziten Verzicht auf eine zwingende, durchgehende Begleitung des Publikums mittels eines körperlich präsenten, virtuellen Gesprächspartners.

³ Das Bild zeigt eine Mixed Reality Anwendung des ZGDV e.V., vgl. Braun [Bra02a]. Mixed Reality bezeichnet eine Vermischung von virtuellen und realen Ein/Ausgabemodalitäten.

Im Gegensatz zu Windows-basierten Systemen, die ebenfalls unter Verzicht von anthropomorphen Gesprächspartnern einen gewissen konversationalen Grad der Interaktion mit dem Publikum erreichen, wird jedoch die Konversation selbst menschähnlich geführt. Dies bedeutet unter Anderem das Aufgeben der Kontrolle der Gesprächsführung durch die Maschine⁴. Die Konversation wird erst unter gleichberechtigten Partnern menschähnlich, dies bedeutet dass ein Unterbrechen eines Konversationspartners oder ein Rangverhältnis in der Konversation explizit erwünscht ist. Zu diesem Zweck ist ein regelbasiertes Verhalten des Systems geplant, das die Konversation interpretiert und entsprechende menschähnliche Reaktionen des Systems⁵ ermöglicht.

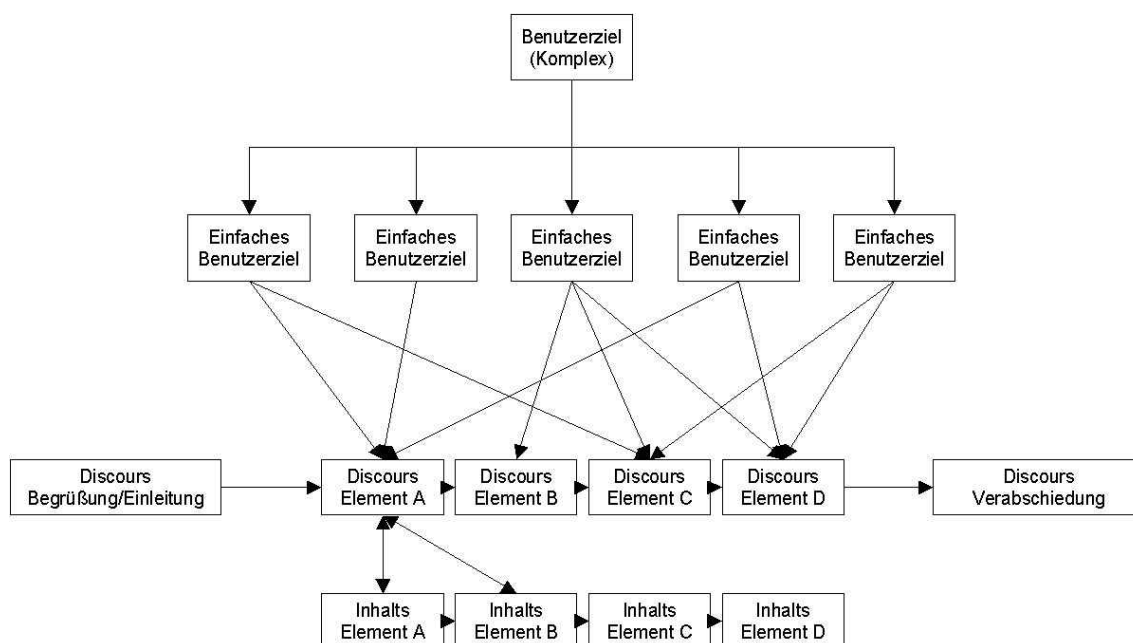


Fig. 4.3: Diskursmodell

Die Diskurs-bezogenen Komponenten können separat in einem Diskurs-Modell notiert werden. Ein Diskurs-Modell beschreibt u.a. die temporalen Eigenschaften zeitbasierter Interaktion. Somit beschreiben Diskurs-Modelle den Ablauf der Konversation aus einer abstrakten, erzählenden Sicht, siehe Abbildung 4.3. Die Abbildung zeigt unter Anderem, dass verschiedene Benutzerziele durch Kombination von präpositionalen (inhaltsbezogenen) und diskurs-bezogenen Elementen gemeinsam bearbeitet werden können. Typische Diskurselemente sind Eröffnung einer Konversation, Turn taking, Feedback-Bereitstellung, Kontrast und Emphatie, Beendigung einer Konversation. Diese Diskurskomponenten können auf mannigfaltige Art und Weise dargestellt werden - so kann zum Beispiel ein Turn taking durch das Heben der Hand, ein Kopfnicken oder ein Räuspern durchgeführt werden, vgl. [CTP98].

⁴ Unter Windows zum Beispiel realisiert durch sogenannte modale Fenster, welche ein explizites turn-taking in der Konversation zwischen Benutzer und System ausformulieren.

⁵ So wird u.a. ein implizites turn-taking realisiert, das dem Benutzer nicht unter den Zwang der Antwort setzt und ihm jederzeit die Möglichkeit des Unterbrechens bzw. Themenwechsels lässt.

Neben den diskursbezogenen Besonderheiten des Konversationsmodells wird eine kombinierte sowohl implizite als auch explizite virtuelle körperliche Repräsentanz des Systems geschaffen: Ist der anthropomorphe Gesprächspartner nicht sichtbar oder hörbar, so ist er doch implizit im Dialog für den Benutzer zu erkennen. Hierzu werden psychologische Modelle menschlichen Verhaltens bemüht, die es erlauben, die Art und Weise der Kommunikation auf die dem Publikum bekannten Eigenheiten der konversationalen Schnittstelle zu projizieren.

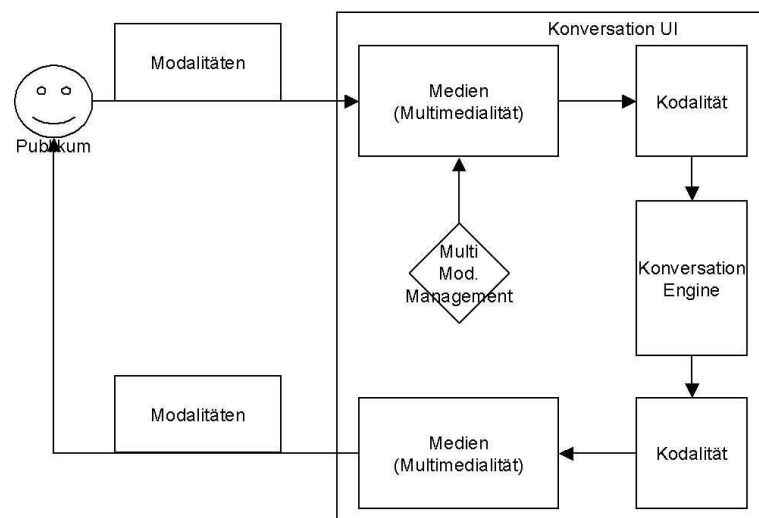


Fig. 4.4: Konversationsbasierte UI Struktur

Die konversationale Interaktion innerhalb einer 3D Umgebung beinhaltet zumeist eine multimodale Komponente. Dies folgt direkt aus der eventuellen Nutzung von Sprache durch den Benutzer und der Präsenz des Benutzers im virtuellen Raum. Durch die gemeinsame Nutzung mehrerer Eingabekanäle zur Bearbeitung einer Aufgabe kann die Fehlerrate der konversationalen Interaktion effektiv gesenkt werden, vgl. [Bol80]. Die Nutzung von mehreren Ein-/Ausgabemodalitäten zur konversationalen Interaktion impliziert die Nutzung von multimodalen Techniken. Die Verwendung dieser multimodalen Techniken geschieht dabei nach dem Prinzip der Redundanz. Redundante Information, die der Benutzer dem konversationalen System zeigt, wird dabei gemappt und zur gezielteren Bestimmung der Informationsinhalte, die der Benutzer an das System weitergibt, bestimmt. Der Benutzer gibt seine Informationen kodiert über verschiedene Modalitäten an das System weiter. Das System muss aus den Medien, die die Modalitäten abbilden, die entsprechende Kodierung herausfiltern, entsprechende konversationale Bausteine kombinieren, kodieren und über die systemeigenen Medienschnittstellen darstellen. Der Benutzer selbst erfährt diese Mediendarstellung als modale Präsentation, siehe Abbildung 4.4.

Eine Modalität ist dabei definiert als ein Reiz, welcher visuell, auditiv, taktil, olfaktorisch, gustativ oder vestibulär ist, siehe Abbildung 4.5. In der Informatik wird Multimodalität oftmals im Zusammenhang mit Multikodalität genutzt. Die Kodalität gibt dabei die Kodierung einer Information an. Multimodale Systeme versuchen, verschiedene Kodalitäten

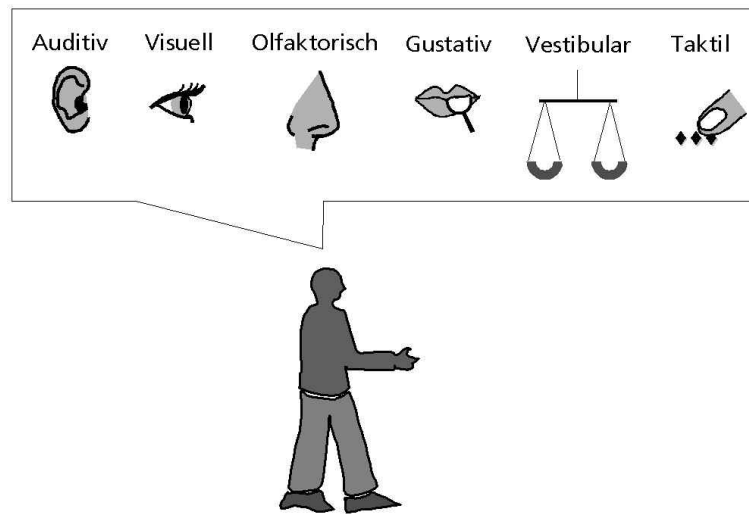


Fig. 4.5: Modalitätskanäle des Benutzers

einer Information zur Zuordnung der Information zu nutzen.

Der Begriff *Multimodalität* wird in diesem Zusammenhang für Ein- und Ausgabe definiert:

- **Multimodale Eingabe:** Unter diesem Begriff wird die Koppelung mehrerer Informationseingabekanäle, wie zum Beispiel gesprochene Sprache (in der Form Audio oder Text), Gestiken (zum Beispiel innerhalb eines Videostreams) und Mimiken (zum Beispiel innerhalb eines Videostreams oder mit sogenannten Emoticons (Icons, welche eine Mimik angeben)), verstanden. Die gekoppelte Eingabe dient der Erreichung eines Zieles. Dieses Ziel kann zum Beispiel eine propositionale (aussagebezogene) Interaktion oder eine diskursbezogene Interaktion sein. An dieser Stelle ist der Unterschied zur multimedialen Eingabe besonders deutlich: Die Nutzung mehrere Medien zur Eingabe (zum Beispiel Audio, Video) ist nicht unbedingt Voraussetzung für die Nutzung mehrerer Informationskanäle. In einem Videostream können 2 Informationskanäle, zum Beispiel Mimik und Gestik, kodiert sein.
- **Multimodale Ausgabe:** Die Koppelung mehrerer Informationsdarstellungen zur Ausgabe einer Information, wie die Kombination von Sprache, Mimik und Gestik, wird unter diesem Begriff verstanden. Beim Begriff der Multimodalen Eingabe ist es auch hier keine notwendige Voraussetzung, mehrere Medien zur Darstellung der Information zu nutzen. Mimik und Gestik, die bei der Multimodalen Ausgabe als getrennte Informationsdarbietungen verstanden werden, können im gleichen Videostream eingebettet sein.

Die Nutzung von multimodaler Ein- und Ausgabe steht in direktem Zusammenhang mit der Nutzung von Körpersprache als zusätzlichem notwendigem Attribut der Konversation. Körpersprache teilt sich ein in Mimik und Gestik. Mimik bezeichnet die (emotionalen) Gesichtszüge einer Person, Gestik bezeichnet die Haltung und die körperlichen Aktionen einer

Person. Mimik und Gestik werden während einer zwischenmenschlichen Kommunikation dauernd - wenn auch zumeist unbewusst - eingesetzt.

Eine beispielhafte Analyse ist im Folgenden dargestellt - sie dienen u.a. zur Akzeptanzsteigerung und Erwartungskonformität seitens der Benutzer. Zu diesem Zweck wurden zwischenmenschliche Ausdrucksweisen untersucht und bewertet, dies in Bezug auf Mimikvielfältigkeit und deren Darstellungsmöglichkeiten.



Fig. 4.6: Benutzer zeigt Interesse an Thema



Fig. 4.7: Benutzer zeigt Interesse an Thema

Eine Liste von Intensionen von Benutzern wurde erstellt und verschiedenen Personen als Vorgabe für entsprechendes Verhalten angewiesen, vgl. [BGS01a]. Die verschiedenen Mimiken der jeweiligen Person wurden auf Video aufgenommen und als jeweils einzelner Video-Clip (siehe Abbildung 4.6) archiviert. Auf diese Weise war es möglich, differenzierte Verhaltensintensionen der Benutzer aufzunehmen und auf ihren sichtbaren Effekt zu bewerten. Diese Aufnahme bildete die Grundlage zur Analyse der Mimikvielfältigkeit und deren



Fig. 4.8: Benutzer zeigt Interesse an Thema

Darstellungsmöglichkeiten im MAP-Kontext.

Weiterhin wurden die Videoaufnahmen im Hinblick auf die verwendeten Endgeräte in mehreren Stufen vereinfacht. Verschiedene Standbilder sind hierfür in der ersten Stufe von einem Designer in Zeichnungen (Abbildung 4.7) transferiert worden. Diese wiederum bilden die Grundlage zur Modellierung verschiedener 3D-Assistenzmimiken (Abbildung 4.8) in relativ hoher Auflösung.

Die modellierten Mimiken und komplexere Verhaltensweisen wurden in speziellen Szenarios getestet, vgl. [BGS01b]. So sind speziell für rein graphische (Abbildung 4.9), rein akustische (Abbildung 4.10) und graphisch akustische (Abbildung 4.11) Szenarios entsprechende Methoden der konversationalen Interaktion simuliert worden, um die in Kapitel 3.1.2 beschriebenen Kameratechniken zur Umsetzung von konversationalen Anforderungen zu prüfen.

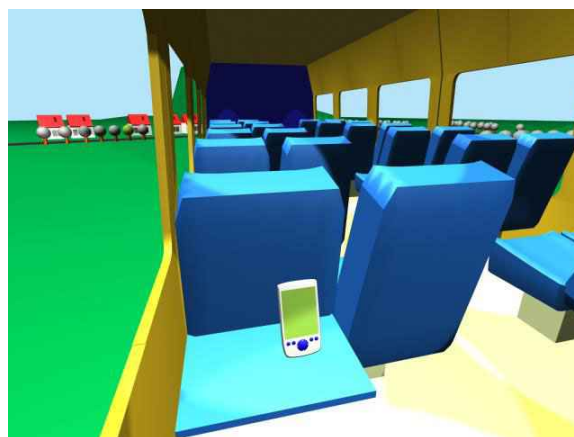


Fig. 4.9: Zug-Szenario zur Simulation graphisch-basierter konversationaler Interaktion



Fig. 4.10: Auto-Szenario zur Simulation akustisch-basierter konversationaler Interaktion

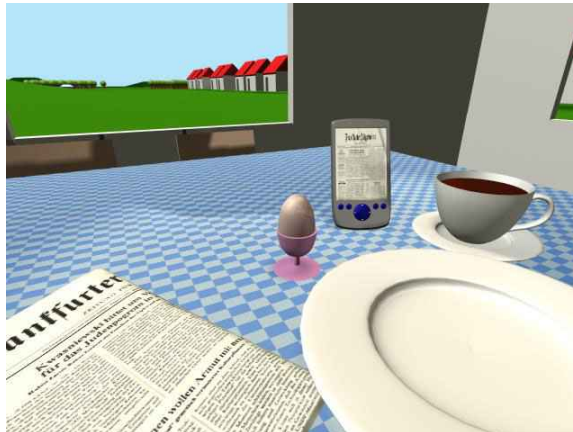


Fig. 4.11: Haus-Szenario zur Simulation graphisch-akustisch-basierter konversationaler Interaktion

Die Analyse von Videoaufnahmen, Bilddokumenten und Literatur (vgl. [Mol91]) über die psychologischen und sozialen Aspekte von Konversation führt zu folgenden Faktoren (die Liste kann beliebig erweitert werden. Um die Nutzbarkeit der Faktoren und damit des Modells zu gewährleisten, sind nur die wichtigsten Faktoren gelistet, unterschieden nach propositionalem oder diskurs-bezogenem Aspekt durch spezielle Markierung der propositionalen Aspekte):

- Soziale und emotionale Aspekte wie Hierarchien, und Beziehungen.
- Geschichte: dramaturgische Inhalts-Sequenzen, asynchrone Daten.
- Immersion: Möglichkeiten der Störung, z.B. im Falle von Assistenz oder interaktiven Filmen.
- Fokus: Die aktuelle Aufmerksamkeit der Benutzer - sieht der Benutzer zum Avatar

oder etwa durch die Frontscheibe seines Autos?

- Konkrete Inhalte (propositional): Tatsächlich bezieht sich dieser Aspekt auf Meta-Informationen über Inhalte. Ist der auszugebende Inhalt eine Frage, Antwort oder ein simpler Ausdruck, welche Bezüge haben die Inhalte untereinander, von wem stammen die Inhalte, d.h. vom Benutzer oder vom System? Weiterhin wird die Dringlichkeit und die Wichtigkeit eines Inhaltes sowie seine emotionale Einschätzung - d.h. die Unterscheidung in gute oder schlechte Nachricht - notiert.
- Navigatorische Aspekte wie Eröffnung und Abschließen eines Diskurses, Rederecht holen oder abgeben.

Offensichtlich sind diese konversationalen Aspekte abstrakt-symbolisch mit einem minimierten Bezug auf den Content, jedoch mit einem maximierten Bezug auf die allgemeinen Verhaltensweisen eines Menschen oder menschähnlichen Konversationspartners. Insofern sind die Aspekte ohne einen modalen Bezug - die modale Ausprägung wird in den entsprechenden Ausgabe- bzw. Eingabemodulen hergestellt, vgl. [Bra01b].

Neben dem minimalen Bezug zum präsentierten Inhalt besteht des Weiteren nur eine geringe Verwandtschaft zu den sogenannten Performatives⁶: Performatives beschreiben die zwischen Agenten zulässigen Operationen, um auf das jeweilige Wissen der anderen Agenten zugreifen zu können und diesen die eigenen Ziele mitzuteilen. Performatives beschreiben insofern nicht das 'zwischenmenschliche' Verhalten von Agenten, da dieses in der Agentenkommunikation nicht relevant ist und keine Beachtung findet.

Auch sind Speech Acts nur sehr lose verwandt mit konversationalen Aspekten. Speech Acts beschreiben den semantischen Kontext von gesprochenen Sätzen, während konversationale Aspekte die allgemeinen Verhaltensweisen zwischen 2 Konversationsteilnehmern während einer Konversation beschreiben.

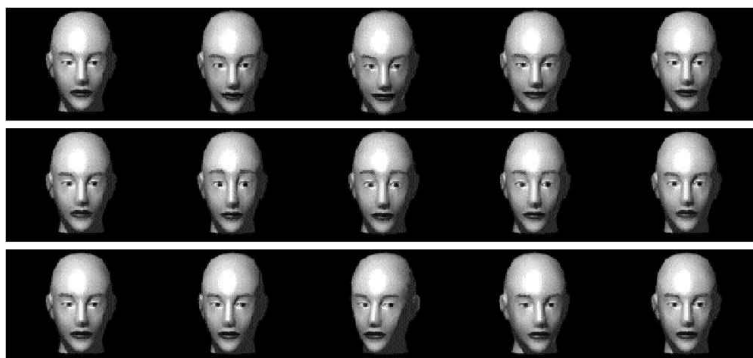


Fig. 4.12: Mimikvarianten eines virtuellen Ansprechpartners

⁶ Performatives sind eine aus der Agentenkommunikationssprache KQML, vgl. [LF94], bekannte Art der Semantikzuordnung der Agentenkommunikation.



Fig. 4.13: Gestikvarianten eines virtuellen Ansprechpartners

Die Abbildung 4.12 zeigt in der oberen Bildhälfte eine bejahende Mimik, in der mittleren Bildhälfte eine Mimik des nicht Verstehens und in der unteren Bildhälfte eine verneinende Mimik. Die Abbildung 4.13 zeigt von rechts nach links folgende Gestiken: Gesprächseröffnung, Turntaking zum Gesprächspartner, Reden, Zeigen, Erklären, Zeigen.

In der Literatur sind diverse konversationale Ansätze beschrieben, die einen Top-Down Ansatz der Modellierung vorgeben. Die Arbeiten von Ball, vgl. [BLK⁺97], sehen einen Gesprächspartner in Form eines Papageis vor. Leider nutzt diese Arbeit keine multimodale Ausgabe zur Konversation zwischen Benutzer und Avatar, ebenso nutzt sie nicht die Verhaltensweisen des Benutzers zur Regulierung der Konversation. Bates und Loyall, vgl. [LB97], nutzen ebenfalls keinen Avatar mit menschlichen Zügen. Der Avatar hat jedoch ein sehr natürliches Verhalten welches Benutzern erlaubt, über längere Zeit mit dem Avatar in Interaktion zu treten. Badler, vgl. [NB97], kombiniert Verhaltensregeln mit Sprachausgabe, erzeugt diese Verhaltensregeln jedoch nicht automatisch. Thorisson, vgl. [TC96], kombiniert Diskurs-Funktionen mit nonverbalem Verhalten als Grundlage eines multimodalen Interfaces mit einem Avatar, welcher über menschliche Gestik und Mimik verfügt. Allerdings erlaubt der entsprechende Avatar nur schwerlich den Einsatz als generativer, autorenbasierter bzw. Story Engine-automatisierten Teil einer Videopräsentation. Der Ansatz von Billinghurst, vgl. [CBB⁺99], nutzt zwar einen virtuellen, körperlichen Avatar als Repräsentanten des Systems, dieser verfügt jedoch über kein hierarchisch organisiertes Verhalten, das mittels menschähnlicher Mimik und Gestik dargestellt wird.

4.1.3 Zusammenfassung

Nach der Gegenüberstellung von Top-Down und Bottom-Up Ansätzen lässt sich die Schlussfolgerung ziehen, dass mittels eines Bottom-Up getriebenen Ansatzes die wichtigen zwischenmenschlichen Konversationsaspekte, die eine kontinuierliche Ausprägung besitzen und im Wesentlichen amodal beschrieben werden können, nicht ausreichend berücksichtigt werden. Leider kann von einer ausreichenden Berücksichtigung der amodalen Aspekte, wie sie in den Top-Down Ansätzen berücksichtigt werden sollten auch nicht gesprochen werden.

Die explizite Modellierung von amodalen, konversationalen Verhaltensweisen ist somit eine Forderung, die zur Regulierung der konversationalen Modellierung unbedingt erfüllt werden

muss. Die amodale Komponente ist weitgehend unabhängig von tatsächlichen Inhalten, die modal präsentiert werden sollten. Die modale Umsetzung sollte in einer zweiten Komponente durchgeführt werden und auf die symbolische Modellierung folgen. Somit kann sie als personifizierter Konversationspartner des Benutzers dienen.

Der in dieser Arbeit verfolgte Ansatz kombiniert einen expertenbasierten Ansatz (Verhalten und Text parallel zu einer Videopräsentation) mit einer automatischen konversationalen Verhaltensgenerierung welche tatsächlich die amodalen Komponenten der Konversation, basierend auf den aktuellen Benutzerinteraktionen, vgl. [BF00], symbolisch darstellt⁷. Er legt dabei seinen Schwerpunkt vor Allem auf menschähnliches konversationales Verhalten sowie auf eine extreme menschähnliche Darstellung des Avatars sowohl in Bezug auf Mimik als auch auf Gestik zur modalen Umsetzung dieses Verhaltens.

4.2 Akustische Informationsdarstellung

Die akustische Informationsdarstellung mit nicht sprachlichen Mitteln wurde bisher vor allem als Erweiterung von GUI (Graphical User Interfaces) und als Hilfe für Sehbehinderte betrieben. Das Hauptziel bei diesen Bestrebungen ist i.A. die akustische Darstellung von diskreter Information, z.B. Systemzuständen, Benutzerinteraktionen, Systemfeedback. In wenigen Fällen wird akustische Informationsdarstellung auch interaktiv genutzt, d.h. zum Beispiel zur Auswahl von akustischen Objekten in einer akustischen Umgebung (das *browsen* von Audiodaten) oder beim Anzeigen von Hyperlinks in akustischen Dokumenten. Neben diesen Möglichkeiten besteht noch die Nutzung des Audiokanals als rein akustischer Zugang zu Computersystemen - zumeist aber beschränkt auf die Darstellung diskreter Information. Im weiteren wird eine Sequenz akustischer Daten auch *Sound* genannt. Eine Übersicht zu akustischer Informationsdarstellung, speziell von Variationsmöglichkeiten für Ton, gibt u.a. Astheimer, vgl. [Ast98].

4.2.1 Akustische Erweiterungen von GUI

Der Motivation von Systemen, die akustische Erweiterungen an GUIs vornehmen, findet sich zumeist in folgenden Punkten:

- Reduzierung der graphischen Komponenten der GUI, um diese übersichtlicher zu gestalten. Verschiedene GUI-Komponenten werden durch akustische Signale dargestellt, z.B.:
 - Status/Überwachungsmeldungen wie *Neue Nachricht*
 - Alarm/Warnmeldungen wie *Batterie fast leer*
 - Kodierte Nachrichten/Daten wie *Einfahrt einer Lokomotive in einen simulierten Bahnhof*

⁷ Vgl. Kapitel 5.2.2.

Empirische Untersuchungen wurden unter anderem von Brown, vgl. [BNG89], durchgeführt.

- Redundanz der graphischen Komponenten, um diese besser im Bewusstsein der Benutzer zu platzieren. So sind von Brewster, vgl. [BC97], Tool-Paletten unter Verwendung von akustischen Signalen aufgewertet worden.

Nichtsprachliche akustische Signale werden dabei in zwei Kategorien unterteilt, *Earcons* und *Auditory Icons*.

Earcons werden akustische Signale genannt, die nicht auf ein natürliches Geräusch zurückgehen. So ist zum Beispiel der Standard-Ton der Microsoft-Betriebssysteme (der sogenannte *Ding-Sound*) ein Earcon. Eine voreingenommene Hörweise durch den Benutzer ist im Falle eines Earcons nicht gegeben, da der Ton in einer natürlichen Umgebung nicht vorkommt.

Auditory Icons gehen auf natürliche, dem Benutzer bekannte Geräusche zurück. Hier wird die Voreingenommenheit des Benutzers gegen diese Töne ausgenutzt, um ein intuitives Verstehen des Tones durch den Benutzer zu erleichtern. So wird zum Beispiel der Ton einer sich schließenden Tür oftmals zum Vertonen einer sich im Moment schließenden Applikation benutzt.

Beispiele für den exzessiven Einsatz von akustischen Objekten in Benutzeroberflächen finden sich bei Gaver, vgl. [Gav89]. Dort werden zwei Sorten von akustischen Objekten benutzt: Einmal Objekte, welche ertönen, sobald ein Ereignis eintritt (Sonic Finder). Zum Zweiten werden Objekte benutzt, die permanent einen Laut von sich geben (Soundholder). Je weiter der Benutzer mit dem Cursor von dem Objekt entfernt ist, desto leiser wird der Laut.

Tatsächlich ist durch Buxton, vgl. [Bux95], nachgewiesen, dass Sound den Immersionsfaktor von Software steigert. Jedoch genügt es nicht, schriftliche Systemmeldungen mit Sound zu unterlegen. Vielmehr bedarf es eines Gesamtkonzepts, das Sound und Graphic schon bei der Entwicklung des UI berücksichtigt. Für akustische Erweiterungen von Systemen wie graphischen GUIs ist insgesamt zu sagen, dass eine Hauptanforderung für akustische Signale eine Deaktivierungsmöglichkeit dieser Signale ist.

4.2.2 Akustisches Browsen von Audiodaten

Das Durchsuchen von akustischen Daten ist wegen ihrer temporalen Natur mit der selben Problematik behaftet wie das Suchen von Daten innerhalb eines jeden kontinuierlichen Mediums. Die Daten befinden sich an einer bestimmten Stelle innerhalb der zeitlichen Ausdehnung im kontinuierlichen Medium. Um sie einzusehen, muss an die entsprechende Stelle des Mediums gewechselt werden. Bezüglich akustischer Daten wird jedoch eine Eigenart des menschlichen Hörens genutzt: Der Cocktail-Party-Effekt, vgl. [Aro92]. Dieser Effekt bewirkt, dass verschiedene akustische Quellen in einer spatialen Umgebung tatsächlich nicht

als ein Geräusch, sondern als mehrere Geräusche unterschieden werden. Durch die räumliche Darstellung des Sounds, d.h. des Anspielens der entsprechenden Töne in verschiedenen spatialen Bereichen eines Raumes, kann einem Benutzer ein Eindruck des vorhandenen Sounds gewährt werden, ohne dass ein direktes Anspielen verschiedener Teile des Sounds nötig sind. Eine Lösung in diesem Bereich wurde unter anderem von Kobayashi, vgl. [KS97], vorgestellt. Dort wird ein einziges Musikstück auf verschiedene spatiale Positionen im Raum verteilt. Jeweils ein anderer zeitlicher Teil des Musikstücks wird an den Positionen abgespielt. Schmandt, vgl. [SM95], hat die selbe Technik zur Wiedergabe von verschiedenen Musikstücken auf verschiedenen spatialen Positionen im Raum benutzt.

4.2.3 Akustische Annotation von Audiodaten

Neben dem Durchsuchen von akustischen Daten ist deren Annotation eine interessante Herausforderung. Annotationen von Audiodaten sind ebenso zeitbehaftet wie die Audiodaten selbst. Dies bedeutet, dass eine Annotation dem Benutzer nicht dauerhaft zur Verfügung steht, jede Annotation besitzt einen Anfang, eine Dauer und ein Ende. Die Annotation selbst kann im Audio-Medium *intramedial*, d.h. innerhalb des Audiokanals, aber auch parallel zum Audio-Medium in anderen Medien *intermedial* (z.B. graphisch) platziert werden⁸. An dieser Stelle sollen nur die Annotationen innerhalb des Audiokanals diskutiert werden. Ein solcher Ansatz wird unter anderem durch Sawhney, vgl. [SM96], verfolgt. Dort werden akustische Ströme temporal mit akustischen Signalen indiziert, welche eine Annotation darstellen und einen Hyperlink zu einem weiteren akustischen Strom signalisieren.

Die akustische Annotation von Audiodaten zur Darstellung von Hyperlinks wurde von Braun, vgl. [BD98a, BD98b], erweitert. Akustische Annotationen im Audiokanal eines Videos werden verwendet zur Verlinkung von akustischen Daten mit URL⁹-referenzierbaren Daten wie Text, Bild, Ton, Animation und Video. Die Interaktion auf das akustische Signal wird mittels Sprachinteraktion auf den akustischen Kanal verlegt. Eigenschaften von akustischen Hyperlinks sind zum Beispiel folgende:

- Sensitive Zeit: Die akustische Darstellung einer akustischen Annotation ist in der Regel kürzer als die tatsächlich gewährte Interaktionszeit, um dem Benutzer eine Reaktion auf den Schluß einer akustischen Annotation zu gewähren, siehe Abbildung 4.14.
- Sound: Der Sound der Annotation. Dieser kann - je nach Anwendung und Intention des Autoren - künstlichen Ursprungs sein (Earcon) oder eine natürliche Quelle besitzen (Auditory Icon).
- Link-Ziel: Das Link-Ziel ist die Medien-Quelle, auf welche der akustische Hyperlink verzweigt.

⁸ Diese Thematik ist in Kapitel 4.3.2.1 vertieft.

⁹ Ein URL (uniform resource locator) ist ein Character String, welcher die Zugriffsmethode und die Lokation einer Ressource im Internet beschreibt.

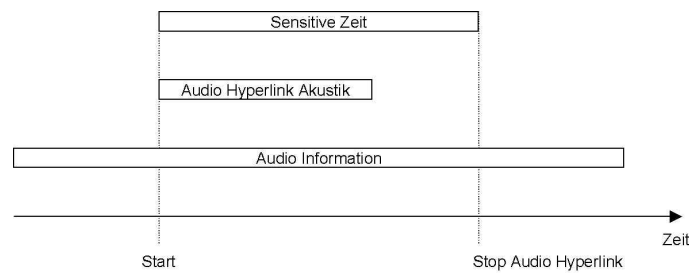


Fig. 4.14: Sensitive Zeit einer akustischen Annotation

- Navigationsmetapher: Die Art, wie ein Benutzer auf die Annotation reagieren kann, z.B. mittels eines Buttons (WIMP) oder mittels eines Sprachbefehls (Conversational, vgl. [Bra99])
- System Reaktion: Da die Interaktion des Benutzers auf einem kontinuierlichen Medium stattfindet, sind verschiedene Systemreaktionen denkbar. So ist - je nach Intention des Autors einer Präsentation - für einen nicht innerhalb des Mediums angesiedelten Link zur Navigation das Stoppen oder das Fortfahren der aktuellen Medienpräsentation eine im Einzelfall zu entscheidende Alternative. Findet eine Navigation innerhalb des Mediums statt, so kann das Springen zur entsprechenden Stelle im Medium oder ein weiteres Öffnen einer Präsentation des gleichen Mediums sinnvoll sein.

Dies führte zu einer intuitiven, intramedialen Interaktionsform auf dem Medium Audio.

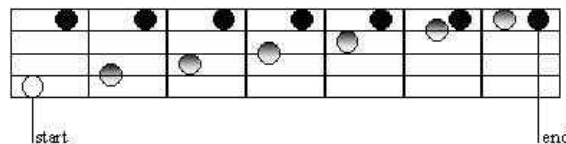


Fig. 4.15: Temporale akustische Annotation von akustischen Daten

Die Problematik der Kontinuität von Audio, d.h. der zeitlichen Ausdehnung einer entsprechenden akustischen Annotation, wurde von Braun, vgl. [BD99], beschrieben. Durch eine intuitive Form der akustischen Darstellung der Annotation kann erreicht werden, dass sofort nach dem Start der akustischen Annotation dem Benutzer angezeigt wird, wie lange die Annotation zugreifbar ist. Somit wird der Stress im Umgang mit akustischen Annotationen für den Benutzer auf ein minimales Maß reduziert. Der Ansatz wird in Abbildung 4.15 mittels einer symbolischen Notation gezeigt und in Kapitel 4.3.2.1 im Zusammenhang mit temporaler Videoannotation vertieft.

4.2.4 Akustische UI

In rein akustischen Benutzerschnittstellen wird - etwa weil die Benutzer sehbehindert sind oder weil kein Display zur Verfügung steht - auf jegliche graphische Ausgabe von Informa-

tion verzichtet. Alle Informationen, d.h. auch Informationen zum Steuern des Gerätes bzw. des Systems, werden akustisch ausgegeben. Ein Großteil dieser Systeme nutzt außerdem den Audiokanal zur Eingabe¹⁰ von Informationen durch den Benutzer.

An dieser Stelle sollen zwei Beispiele angegeben werden. So wurde von Boelke, vgl. [BG95], ein System vorgestellt, welches blinden Benutzern ermöglicht, durch direkte Manipulation von akustischen Objekten das System zu nutzen. Es verwendet sogenannte Hearcons, welche durch folgende Eigenschaften charakterisiert werden:

- Ein Geräusch, welches das Objekt repräsentiert.
- Die Lautstärke des Geräusches.
- Die spatiale Position des Geräusches.
- Die spatiale Ausdehnung des Geräusches.

In der Menge der Hearcons befindet sich ein ausgezeichnetes Hearcon, das aktuelle Hearcon. Nur von diesem Hearcon kann der Benutzer Informationen beziehen. Durch diesen Ansatz ist es dem Benutzer möglich, ein Objekt gezielt zu selektieren und zu manipulieren. Durch die Verwendung von einfachen Geräuschen wird in diesem Ansatz die Problematik der temporalen Ausdehnung von akustischen Informationen weitgehend umgangen.

Stiefelmann, vgl. [Sti95, Sti96], stellt eine reine akustische Schnittstelle zu Benutzern vor, welche Sprachinformationen zur Systemsteuerung und zum Systemfeedback nutzt. Die Applikation wird zum akustischen Annotieren von Aufzeichnungen auf Papier benutzt.

4.2.5 Zusammenfassung

Der interaktive Zugang zu akustischen Daten wird geprägt durch die temporale Ausdehnung des Mediums Audio. Von den gezeigten Ansätzen arbeitet nur die akustische Annotation von Audio (Kapitel 4.2.3) als echte interaktive Schnittstelle zu akustischen Daten. In allen anderen gezeigten Bereichen wird die akustische Darstellung zwar als Ereignis/Objekt-Anzeige benutzt, es werden jedoch keine wirklichen kodierten Nachrichten und Informationen mittels des Audio-Kanals dargestellt. So kann in den entsprechenden Ansätzen auf die akustischen Daten nur als Ganzes zugegriffen werden - d.h. der Benutzer kann nicht mit interessanten Teilen der Audio-Informationen interagieren. Insofern wurde der Audiokanal seiner temporalen Ausdehnung beraubt und nicht mehr als kontinuierliches Medium genutzt.

Die Forderung an dieser Stelle beinhaltet eine Möglichkeit der Audio-Interaktion, welche auf die temporale (und spatiale) Ausprägung des Mediums eingeht und den Benutzer bei der Interaktion effektiv unterstützt. Diese Forderung wird von der akustischen Annotation von Audio erfüllt.

¹⁰ Akustische Eingabe ist im Kapitel 4.1.1 diskutiert.

4.3 Graphische Interaktion auf kontinuierlichen Medien

Die graphische Informationsdarstellung und Interaktion besitzt eine lange Historie. In dieser Arbeit wird ausschließlich auf die Interaktion auf graphisch-kontinuierlichen Medien mit dem Schwerpunkt konversationale Interaktion und Hypermedia eingegangen. Eine umfassende Zusammenfassung zum Thema Interaktion auf graphischen Medien ist unter anderem bei Buxton, vgl. [Bux98], zu finden.

4.3.1 Animation und VR

Die Interaktion in 3D Umgebungen beeinflusst in starkem Maße das Gefühl der Präsenz des Benutzers in der Szene. Dies ist ein nicht unerheblicher Faktor der Immersion in eine Geschichte. Die Präsenz eines Benutzers in eine Szene kann mit Hilfe von Faktoren angegeben werden, den sogenannten *Presence Questionnaire*, von Singer, vgl. [WS98]. Der *Presence Questionnaire* ordnet das Präsenz-Gefühl eines Benutzers folgenden Faktoren zu:

- Ablenkungsfaktoren: Isolation, Selektive Aufmerksamkeit, Interface Bewusstsein.
- Kontrollfaktoren: Grad der Kontrolle, Unmittelbarkeit der Kontrolle, Erwartung von Ereignissen, Art der Kontrolle, Anpassung der physikalischen Umgebung.
- Realismusfaktoren: Szenenrealismus, Konsistenz der VR-Daten mit Real-World Erfahrungen, Aussagekräftigkeit des Erlebnisses.
- Wahrnehmungsfaktoren: Sensormodalität, Reichhaltigkeit der virtuellen Umgebung, multimodale Präsentation, Konsistenz der multimodalen Daten, Bewegungsempfinden, pro-aktive Komponenten der VR-Szene.

Sogenannte weiche Faktoren wie *Erwartung von Ereignissen* scheinen eine wichtige Rolle beim Präsenzgefühl des Benutzers zu spielen. Da auch ein Buch dem Publikum das Gefühl der Immersion in eine Geschichte geben kann, war dies zu erwarten. So wird unter anderem von Blake, vgl. [JND⁺00] beschrieben, dass der *Presence Questionnaire* zwar ausreicht, um eine VR-Umgebung zu kategorisieren. Zur Bewertung des Präsenzgefühls eines Benutzers kann er jedoch nicht herangezogen werden. Interessanterweise wurden Experimente durchgeführt, vgl. [JND⁺00], welche die Benutzer mit einer Umgebung konfrontierten, die sich zwar in den Faktoren des *Presence Questionnaire* unterschied, jedoch keine Unterschiede bezüglich der Narration der durch die VR-Umgebung vermittelten Geschichte aufwies. Als Resultat wurden beide VR-Umgebungen durch das Publikum mit dem gleichen Immersionsgrad erlebt.

Die Interaktion des Benutzers mit einer 3D Umgebungen kann mit 2D-Metaphern, mit auf 3D angepassten 2D-Metaphern oder mit echten 3D Metaphern geführt werden. Die Eingabegeräte des Benutzers können 2D-Eingabegeräte (zum Beispiel eine Maus, ein Joystick/Gamepad) oder 3D-Eingabegeräte (Headtracker, Data-Glove, 3D-Maus) sein.

- Ein Beispiel für eine 2D-Metapher in einer 3D Umgebung ist das WIMP-Interface. Menus können im 3D Raum dargestellt werden, wie vom Desktop eines Windows-Computers her gewohnt - natürlich mit einigen Nachteilen: In der 3D Welt ist eine Platzierung des Menus räumlich. Die gewohnte Platzierung von Menus auf der obersten Ebene der Bildelemente kann sich störend auf die Immersion im 3D Raum auswirken. Die Selektion von Menuelementen wird durch ein 3D-Eingabegerät wie zum Beispiel einer 3D Maus erschwert, da der Maus-Zeiger sich hinter dem Menu befinden kann.
- Eine 2D Metapher, die an einen 3D Raum angepasst ist, kann zum Beispiel ein räumlich dargestelltes Menu sein. Auch hier existiert die Problematik der Selektion des Menus, allerdings passt sich das räumliche Menu besser in eine 3D Umgebung ein.
- Eine 3D Metapher zur Interaktion ist zum Beispiel eine virtuelle Tür oder ein virtuelles Transportmittel (Auto) im Raum. Im Allgemeinen sind diese Metaphern sehr gut an die 3D Umgebung angepasst, allerdings ungewöhnlich für den Benutzer zu bedienen.

Grundsätzlich kann der Benutzer verschiedene Sichtweisen in der 3D Umgebung einnehmen. Hand, vgl. [Han97], unterscheidet zwischen *Exozentrischen* und *Egozentrischen* Techniken.

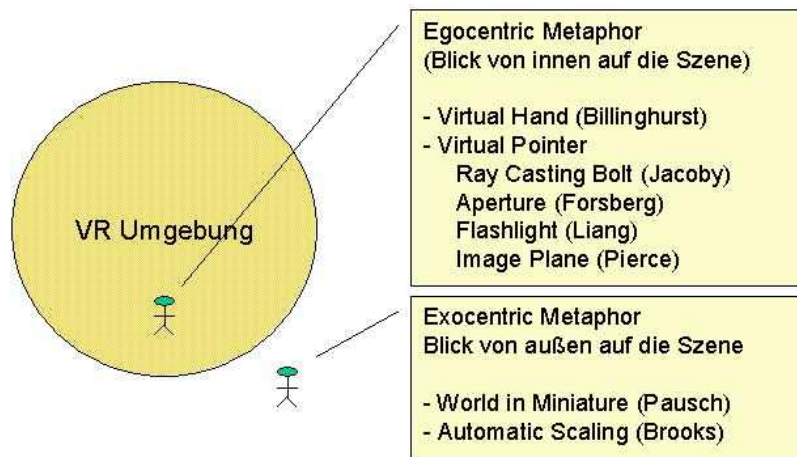


Fig. 4.16: Verschiedene Interaktionsmetaphoriken der VR, vgl. [PWBI98]

Die egozentrischen Techniken, siehe Abbildung 4.16, geben dem Benutzer den Eindruck, sich ständig im Zentrum des VR-Raums zu befinden. Jede Interaktion, zum Beispiel ein Wechsel der virtuellen Lokalität oder der Blickrichtung des Benutzers, wird so durchgeführt, dass der Benutzer die Interaktion *wie durch seine eigenen Augen gesehen* erlebt - quasi von innen aus der Szene. Beispiele für solch einen Interaktionsansatz sind Techniken wie das

sogenannte *Eyeball in Hand*¹¹ und *Flying Vehicle Control*¹².

Im Gegensatz dazu steht die exozentrische Interaktion, siehe Abbildung 4.16. Der Benutzer erlebt die Interaktion nicht durch seine Augen, sondern an einem Objekt, welches er manipuliert. Er erlebt somit die Interaktion quasi von außerhalb der Szene. Ein Beispiel für einen solchen Ansatz ist die *Scene in Hand*¹³ Technik.

4.3.2 Video

Video ist als Medium eine Komposition aus graphischer und akustischer Information. Für Video wird deswegen eine kombinierte, sowohl die Interaktion mit akustischen Daten als auch die Interaktion mit graphischen Daten berücksichtigende Interaktionsmöglichkeit angestrebt.

Video hat gegenüber VR den Nachteil der weniger körperimmersiven Umgebung - das Publikum befindet sich körperlich immer außerhalb der Videoszene. Aus diesem Grunde fallen die egozentrischen graphischen Interaktionsmöglichkeiten der VR für Video weitgehend aus. Das Publikum besitzt immer einen exozentrischen Standpunkt zur Szene.

Die Interaktionen auf dem Medium Video können in 2 Kategorien eingeteilt werden. Aufgrund der temporalen, und damit kontinuierlichen Eigenschaften von Video beziehen sich Interaktionen entweder auf den aktuellen Präsentationszeitpunkt oder zeitunabhängig auf das gesamte Video.

- Die Interaktion auf den aktuellen Präsentationszeitpunkt geschieht quasi synchron zum Ablauf des Videos, sie wird synchrone Interaktion¹⁴ genannt.
- Die zeitunabhängige Interaktion auf das gesamte Video geschieht asynchron zum Ablauf des Videos, sie wird asynchrone Interaktion¹⁵ genannt.

Die verschiedenen Variationen dieser Kategorien werden im Folgenden diskutiert.

4.3.2.1 Synchrone Interaktion auf Video

Die synchrone Interaktion erfolgt mit Bezug auf Informationsobjekte, welche im Video zu sehen oder zu hören sind. Aufgrund der exozentrischen Sichtweise des Publikums geschieht die Interaktion auf den Objekten i.A. über eine Annotation.

¹¹ Eyeball in Hand: Der Benutzer steuert durch eine Szene - der Focus der Szene befindet sich immer in seiner Blickrichtung.

¹² Flying Vehicle Control: Der Benutzer steuert eine Szene mittels einer virtuell dargestellten Steuereinheit, zum Beispiel einem Flugzeugsteuerknüppel.

¹³ Scene in Hand: Der Benutzer manipuliert seinen Standpunkt in der 3D Szene direkt durch Manipulation eines Miniatur-Abbildes der Szene.

¹⁴ Wählt der Benutzer ein annotiertes Objekt im Video aus und wird dadurch ein Sprung zu einer anderen Stelle im Video ausgelöst, so ist diese Operation synchron zur Videopräsentation.

¹⁵ Eine typische Videorecorder-Funktionalität wie schneller Vorlauf des Videos ist eine zur Inhalts-Präsentation des Videos asynchrone Operation des Benutzers.

Eine Annotation bedeutet im ursprünglichen Sinn ausschließlich die Präsenz einer Markierung. Diese Markierung signalisiert dem Publikum eine Interaktionsmöglichkeit - jedoch nicht welche Interaktionsmöglichkeit. Im Folgenden werden spatiale und temporale Annotationen von Informationsobjekten in Video diskutiert, jedoch nicht die Aktionen¹⁶, welche durch sie ausgelöst werden.

Eine Annotation kann sich innerhalb des Mediums befinden (etwa innerhalb der Graphik oder Akustik des Videos), sie kann jedoch auch außerhalb des Videos, quasi in einem separaten Medium erfolgen. Erfolgt die Annotation innerhalb des Mediums, so wird dies eine intramediale Annotation genannt. Die Interaktion auf der Annotation wird dementsprechend eine intramediale Interaktion genannt. Im umgekehrten Fall erfolgt die Annotation außerhalb des Mediums in einem separaten Medium, weshalb die Annotation eine intermediale Annotation genannt wird. Entsprechend wird die Interaktion des Publikums eine intermediale Interaktion genannt.

Die synchrone Interaktion mittels Annotation wird von einem Zeitfaktor beherrscht, vgl. Braun [BD99]. Jedes annotierte Objekt besitzt einen Präsentationsstart, eine Präsentationsdauer und ein Präsentationsende. Die Annotation des Objektes besitzt dementsprechend ebenfalls einen Startzeitpunkt, eine temporale Dauer und einen Endzeitpunkt. Im Allgemeinen werden diese temporalen Ausprägungen jedoch nicht dargestellt. Dies führt zu einer Stresssituation für das Publikum, da kein Indiz vorhanden ist, um die zeitlich begrenzte Interaktionsmöglichkeit während der Dauer der Annotation zu ermessen. Wird die Dauer einer Annotation explizit dargestellt, so wird dies temporale Annotation genannt. Um eine temporale Darstellung der Annotation zu erreichen muss die Annotation folgende Bedingungen erfüllen:

- Die Annotation muss beim Start die gesamte verbleibende Interaktionszeit anzeigen.
- Die Annotation muss während der Dauer der Annotation die verbleibende Interaktionszeit anzeigen.
- Die Annotation muss zum Ende der Annotationszeit explizit das Ende der Annotation anzeigen.

In Abschnitt *Intramediale Interaktion* wird auf die temporale Darstellung von Annotationen weiter eingegangen.

Intermediale Interaktion Die intermediale Interaktion des Publikums auf ein Video erfolgt auf Annotationen, welche außerhalb des Videos in einem separaten Medium dargestellt werden, vgl. Braun [BF00]. So erfolgt zum Beispiel in Hypercafe¹⁷, vgl. [SBS96], die synchrone Auswahl von Folge-Clips mittels Icons, welche eine Szene des Folge-Clips darstellt. Eine weitere Möglichkeit ist das synchrone Einblenden von Web-Seiten, die ihrerseits über Hyperlinks verfügen, siehe Gerfelder [AGGS96]. Eine Verfeinerung dieses Ansatzes ist das

¹⁶ Die ausgelösten Aktionen sind Bestandteil der Ästhetik des nichtlinearen Videos und damit weitgehend dem Autoren einer nichtlinearen Präsentation überlassen. Sie werden in Kapitel 4.4 diskutiert.

¹⁷ Vgl. Kapitel 3.3.1.2.

synchrone Einblenden von Texthyperlinks quasi neben dem Video. Diese Technik wurde von Digital Renaissance, vgl. [Mar96] eingesetzt. Die Problematik des Ansatzes ergibt sich aus der für das Publikum wenig intuitive Zuordnung der Annotation zum Objekt im Video. Da die Annotation nicht direkt mit dem Objekt verknüpft ist, muss eine spezielle Bedeutung in die Namensgebung der Texthyperlinks eingebracht werden, um dem Publikum die Zuordnung von Annotation und annotiertem Objekt zu ermöglichen. Dies führt zu einer zusätzlichen Semantik, die nicht direkt mit der Inhaltswiedergabe des Video in Zusammenhang steht und daher das Publikum unnötig belastet.

Intramediale Interaktion Hier erfolgt die Interaktion des Publikums innerhalb eines Mediums. Dies bedeutet, dass die Informationen und die Annotationen im selben Medium, z.B. einem Video, dargestellt werden. Die Interaktion des Publikums erfolgt auf dem gleichen Medium. Diese Vorstellung ist idealisiert, da das Publikum den Eindruck hat, dass die Interaktion auf dem gleichen Medium erfolgt. Tatsächlich erfolgt die Interaktion über eine für den Benutzer logisch verknüpfte Aktion, die in Bezug auf die Modalität dem Counterpart des Mediums entspricht: Bei Graphik ist dies das Zeigen, bei Akustik ist dies das Sprechen.

- Wird die Graphik eines Videos annotiert, so erfolgt die Annotation durch eine graphische Markierung des Informationsobjektes im Video. Die für den Benutzer logische Interaktion ist ein Mausklick (als Repräsentanz einer Zeigegeste) auf die Annotation und damit auf das Objekt.
- Wird die Akustik eines Videos annotiert, so erfolgt die Annotation durch eine akustische Markierung des Informationsobjektes im Video. Die für den Benutzer logische Interaktion ist eine akustische Antwort auf die Annotation. Diese akustische Antwort erfolgt für den Benutzer quasi im gleichen Medium, d.h. im akustischen Kanal.

Die graphische Annotation eines graphischen Informationsobjektes im Video erfolgt entweder explizit oder implizit.

- Explizit: Mittels eines Polygons oder Kreises, welcher zur Dauer der Annotation (temporal) um das Objekt herum gezeichnet wird (spatial).
- Implizit: Das Objekt wird nicht sichtbar markiert. Der Benutzer muss sein Interesse am Objekt bekunden (z.B. durch ein Selektieren mit der Maus) um angezeigt zu bekommen, ob das Objekt annotiert ist.

Die implizite Annotation ist wenig intuitiv für den Benutzer, da der Sinn der Annotation, nämlich das Anzeigen einer Interaktionsmöglichkeit, dadurch verloren geht. Wird die Annotation explizit ausgeführt, so kann durch eine geschickte Manipulation der Annotation die temporale Ausdehnung der Annotation angezeigt werden. Wird das Polygon in der Farbe verändert, z.B. durch ein Ändern der Farbe vom untersten Punkt des Polygons hin zum obersten Punkt des Polygons oder durch ein grad-weises, kreis-förmiges Ändern der Farbe, so kann der Benutzer aufgrund der Geschwindigkeit der Farbänderung sehr genau erkennen,

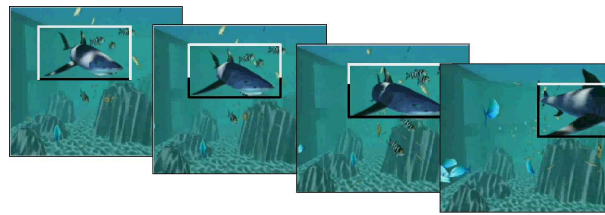


Fig. 4.17: Temporale Darstellung eines Videohyperlink

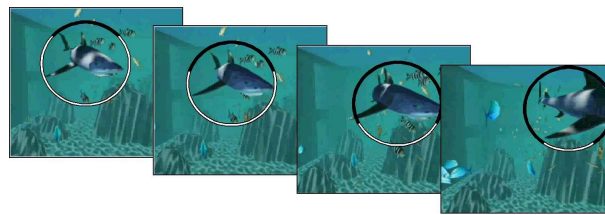


Fig. 4.18: Temporale Darstellung eines Videohyperlink

wann das Polygon die Farbe insgesamt geändert hat und damit die Dauer der Annotation beendet ist, siehe Abbildung 4.17 und 4.18.

Die akustische Annotation eines akustischen Informationsobjektes im Video erfolgt, ähnlich wie bei der graphischen Annotation, ebenfalls explizit oder implizit. Die implizite Variante der Annotation birgt jedoch erhebliche Nachteile: Der Benutzer kann nicht mit einem Mausklick sein Interesse an einer Interaktionsmöglichkeit zum Ausdruck bringen - er muss nachfragen. Dies wird von Benutzern als äußerst störend empfunden. Die explizite Annotation kann durch eine Änderung des Klanges des Objektes, siehe [BD98b], durchgeführt werden oder durch das Abspielen einer zusätzlichen Tonquelle, parallel zum annotierten akustischen Objekt. Dieser zusätzliche Sound kann sowohl ein Auditory Icon als auch ein Earcon sein¹⁸.

Die Verwendung eines Auditory Icon zur akustischen Annotation impliziert Schwierigkeiten bei der Darstellung der temporalen Ausprägung der Annotation: Natürlicher Sound hat eine ihm typische zeitliche Ausdehnung, die schwerlich intuitiv gestreckt oder gestaucht¹⁹ werden kann.

Earcons, da per Definition dem Benutzer vorab nicht aus der Natur bekannt, können sehr einfach dahingehend manipuliert werden, dass sie ihre temporale Ausdehnung tatsächlich für den Benutzer intuitiv verständlich anzeigen. Als Beispiel für eine Earcon-artige Annotation von Sound soll an dieser Stelle der sehr bekannte Microsoft-’Ding’ Sound herangezogen werden. Durch mehrfache Wiederholung des Ding-Sounds in gleichen Abständen entsteht

¹⁸ Vgl. Kapitel 4.2.1.

¹⁹ Als Beispiel soll an dieser Stelle der Sound einer sich schließenden Tür angeführt werden. Die sich schließende Tür mit dem dafür typischen Knarren zeigt dem Benutzer zwar eine temporale Ausdehnung an, jedoch kann er schwerlich abschätzen, wann sich die Tür wirklich schließt und damit eine Annotation dieser Art beendet ist.

ein neuer Sound mit einer gewissen Dauer.

Ein Benutzer kann jedoch die Dauer und das Ende des Sounds nicht abschätzen. Werden die Abstände zwischen den Tönen kontinuierlich verringert (siehe Abbildung 4.19) bzw. die Tondauer selbst verringert (siehe Abbildung 4.20), so hat der Benutzer eine ungefähre Vorstellung von der Dauer des Sounds.



Fig. 4.19: Darstellung der Dauer einer akustischen Annotation mittels Tonwiederholungen



Fig. 4.20: Darstellung der Dauer einer akustischen Annotation mittels Reduzierung der Tondauer

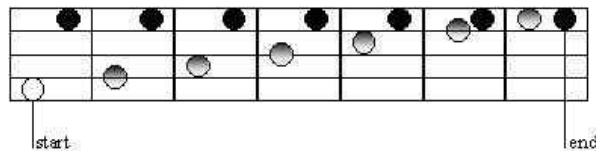


Fig. 4.21: Darstellung einer akustischen Annotation mittels Tonhöhenvariation

Er kann jedoch nicht abschätzen, wann der Sound wirklich zu Ende ist. Werden zwei Tonhöhenvarianten des Ding-Sounds benutzt, so kann schon eine eindeutige Zuordnung des Starts, der Dauer und des Endes einer Annotation durch den Benutzer erfolgen, siehe Abbildung 4.21:

- Ton A wird auf einer Tonhöhe gehalten. Er signalisiert unmittelbar mit dem Start der Annotation das Endsignal der Annotation.
- Ton B wird auf einer niedrigeren Tonhöhe initialisiert. Während der Annotationsdauer wird seine Tonhöhe beständig erhöht - bis er die äquivalente Tonhöhe des Tons A besitzt. Ton B signalisiert dem Benutzer unmittelbar mit dem Start der Annotation die Dauer der Annotation - über den tonhöhen Abstand zu Ton A. Mit der kontinuierlichen Erhöhung der Tonhöhe wird dem Benutzer die verbleibende Dauer der Annotation vermittelt. Besitzen beide Töne A und B die selbe Tonhöhe, so signalisiert dies das Ende der Annotation.

So kann mittels dieser Technik sowohl auf der akustischen als auch auf der graphischen Komponente von Video eine sinnvolle Annotationsmöglichkeit des Mediums angeboten werden.

4.3.2.2 Asynchrone Interaktion auf Video

Die Benutzerinteraktionen, welche asynchron zur Präsentation des Videos erfolgen, können in zwei Kategorien eingeordnet werden, vgl. Braun [BF00].

- Interaktionen, die sich navigatorisch auf die Präsentation des nichtlinearen Videos beziehen. Dies entspricht den diskursbezogenen Interaktionen einer Konversation²⁰.
- Interaktionen, die sich auf den Inhalt des Videos beziehen. Hier wird der aktuelle Präsentationszeitpunkt des Videos tatsächlich als Kontext der Interaktion des Benutzers angesehen. Dies entspricht der präpositionalen Interaktion einer Konversation²¹.

Direkt Manipulative asynchrone Interaktion Die direkte Manipulation eines Videos als Ganzes lässt sich unter dem Begriff *Videorecorderfunktionalität* zusammenfassen. Hier wird das Video als Binary Large Object (Blob) behandelt, entsprechend grob fallen die Interaktionsmöglichkeiten aus: Start, Pause, Vorlauf, Rücklauf sowie Variationen dieser Aktionen.

Eine inhaltsbezogene, direkt manipulative, asynchrone Interaktion kann durch die Visualisierung von Inhalten parallel zur Videopräsentation²² geschehen. Diese Art der Präsentation steht im Gegensatz zur kontinuierlichen Darbietung der Informationen durch ein nichtlineares Video. Der Benutzer wird durch die zusätzliche Informationsaufnahme vom eigentlichen Kontext abgelenkt, was zu einem narrativen Bruch bei der Präsentation des Videos führt.

Konversationale asynchrone Interaktion Die asynchrone Interaktion erfolgt im Wesentlichen, um einen direkten Zugang zu den Informationen zu gewähren, welche durch die kontinuierliche Präsentation des nichtlinearen Videos nur zu bestimmten Zeitpunkten für den Benutzer zugänglich sind. Durch das kontinuierliche Medium Video wird ein Kontext²³ und eine temporale Ausprägung der Konversation erreicht. Dies dient dem Benutzer und dem System als gemeinsame Informations-Grundlage, vgl. Braun [Bra99]. Wird keine virtuelle körperliche Repräsentanz als Ansprechpartner des Systems genutzt, so erweist sich die konversationale Schnittstelle direkt zum Video für den Benutzer aufgrund des fehlenden menschähnlichen Ansprechpartners²⁴ als weniger intuitiv, siehe Abbildung 4.22.

Wichtig - jedoch nicht zwingend, vgl. Braun [Bra00] - für eine intuitiv verlaufende Konversation ist somit eine sichtbare virtuelle Persönlichkeit, die mit dem Benutzer eine gewis-

²⁰ Vgl. Kapitel 4.1.

²¹ Vgl. Kapitel 4.1.

²² Dies wurde unter anderem mit dem System AVWoD (Audio Video Web on Demand), vgl. [AGGS96] gezeigt.

²³ U.a. von Bahrtdt, vgl. [Bah90], gefordert.

²⁴ Krämer und Nitschke, vgl. [KN01] zeigen, dass der Benutzer einen Partner braucht, der offensichtlich ähnliche bis gleiche Aktionsmöglichkeiten verwendet.

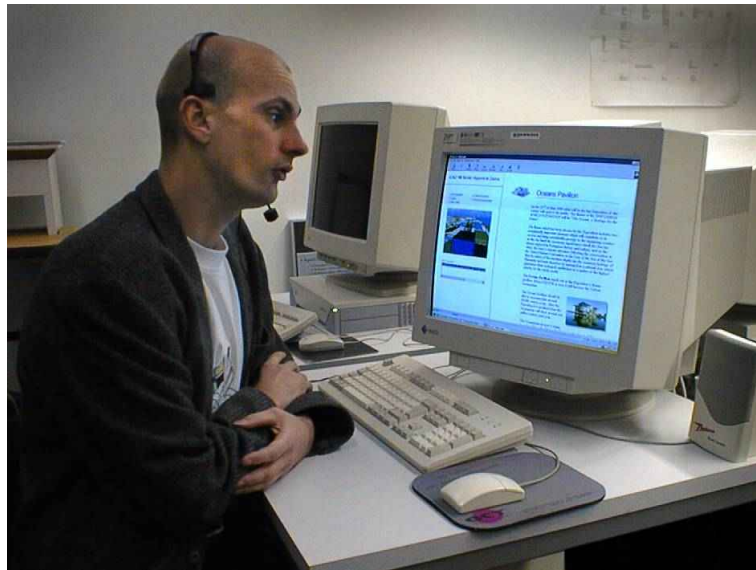


Fig. 4.22: Konversation ohne virtuellen Ansprechpartner

se gemeinsame Grundlage der Kommunikation besitzt, so dass der Benutzer die virtuelle Persönlichkeit tatsächlich als solche erkennt.

4.3.3 Zusammenfassung

Offenbar ist bezüglich der Kontinuität von Medien kein durchgängiges Interaktionsschema zu finden. So werden Versatzstücke der nichtkontinuierlichen Interaktion oftmals auf kontinuierliche Medien übertragen, ohne jedoch auf die zeitlichen Aspekte der Interaktion einzugehen.

Folgende Forderungen werden in dieser Arbeit an die Interaktion mit kontinuierlichen graphischen Medien gestellt: Die zeitlichen Aspekte der Interaktion müssen für den Benutzer auf intuitive Weise erfassbar sein - besser noch ist die Einbringung eines konversationalen Diskurses in die Interaktion mit kontinuierlichen Medien, welche dem Publikum den natürlichen Umgang mit kontinuierlichen Medien extrem erleichtert.

Aus diesem Grund ergibt sich als erweiterte Forderung eine kombinierte Direkt Manipulative und Konversationale Herangehensweise an die Interaktion, um dem Benutzer größtmögliche Freiheit bei gleichzeitig zweckgebundener Interaktionsfähigkeit auf dem Medium zu ermöglichen.

4.4 Autorensysteme für kontinuierliche Medien

Neben der Definition des Begriffs Story Engine dieser Arbeit²⁵, wird der Begriff Story Engine in bezug auf das Authoring von Geschichten in verschiedenen anderen Zusammenhängen benutzt. So wird zum Beispiel ein Werkzeug zum Beschreiben und Erstellen (Autorenschaft) einer Geschichte Story Engine genannt, ebenso wie ein Werkzeug zum Ausführen einer Geschichte Story Engine genannt wird. In die Kategorie Beschreiben und Erstellen können wiederum verschiedene Ansätze entdeckt werden. Einige Werkzeuge beschäftigen sich mit dem amodalen²⁶ Teil (abstrakt-unbildlich) einer Geschichte. Andere Werkzeuge dienen tatsächlich zum modalen²⁷ (gegenständlich-repräsentativ) Erstellen von Szenen einer Geschichte.

Gegenwärtig sind unterschiedliche Autorensysteme für kontinuierliche Medien-Anwendungen auf dem Markt erhältlich, die sich hauptsächlich durch Unterschiede in Leistungsfähigkeit, Bedienphilosophie bzw. Bedienmetaphorik sowie Möglichkeit der Datenbankanbindung auszeichnen. Generell finden Autorensysteme Verwendung bei der Integration unterschiedlicher Dokumente und Mediendateien (Text, Bild, Video, Audio, 3D-Animation) in eine einheitliche, interaktive Anwendung. Neben der Gestaltung und Entwicklung der graphischen Bedienoberfläche (GUI) der Anwendung bieten sie die Möglichkeit, den Ablauf der Anwendung festzulegen, Navigationsmöglichkeiten zu definieren und Interaktion mittels grafischer oder akustischer Elemente einzufügen.

Zu den gebräuchlichsten professionellen Autorensystemen für modales Authoring zählen neben Director 8 Shockwave Studio (Macromedia) und Flash (Macromedia) insbesondere Authorware 5 Attain (Macromedia), Toolbook II (click2learn) Instructor 7.2, Mediator 5 Pro (Matchware). Die genannten Autorensysteme greifen für die weitgehend transparente grafische Darstellung des Entwicklungsprozesses und der strukturellen Bezüge der Sinnzusammenhänge und einzelnen Elemente auf spezielle Metaphern zurück. Je nach der für die Anwendung spezifischen Zielausrichtung des Autorensystems lassen sich Buch-Metapher, Flussdiagramm-Metapher und Film-Metapher unterscheiden. Im Folgenden werden die verschiedenen Authoringmetaphern näher diskutiert:

Buch-Metapher-orientierte Autorensysteme Die Buch-Metapher findet bei dem überwiegenden Teil der auf dem Markt befindlichen Autorensysteme Verwendung (z.B. Toolbook, Mediator 5 Pro, Katabounga 3.0, Hypermethod Professional 3.0). Sie dient sowohl zur Organisation der inhaltlichen Bezüge und Zusammenhänge, als auch zur grafischen Darstellung des Entwicklungsprozesses der jeweiligen multimedialen Anwendung bzw. Präsentation. Die zu entwickelnde Multimedia-Anwendung stellt sich hierbei als ein aus einzelnen interaktiven Bildschirmseiten bestehendes Buch dar, das in der Endanwendung von den Usern interaktiv durchgeblättert werden kann. Der Vorteil der Buch-Metapher-orientierten Autorensysteme liegt darin, dass bei der Entwicklung einer Anwendung alle Inhalts-Seiten stark verkleinert nebeneinander betrachtet

²⁵ Vgl. Kapitel 3.2.1.

²⁶ Daher mit der Geschichtsstruktur, mit den beteiligten Charakteren, den Aktionen.

²⁷ Diese modellieren Aussehen, Ambiente, Texte etc..

werden können und ein relativ guter Überblick über das vorhandene Material und die zum Einsatz kommenden Medien möglich ist. Der Nachteil dieser Systeme ist in der spezifischen, graphischen Repräsentation der Buch-Metapher begründet - der Programmablauf sowie dessen durchaus komplexe Verzweigungsoptionen lassen sich nur sehr eingeschränkt bzw. gar nicht visualisieren. Somit sind Buch-Metapher-orientierte Autorensysteme nur mit erheblichen Einschränkungen für die konzeptionelle Entwicklung, die Gestaltung und das Management komplexer Szenarios auf der Grundlage unterschiedlichster Verzweigungsoptionen und Use Cases einsetzbar. Hierfür scheinen auf den ersten Blick Autorensysteme besser geeignet zu sein, die in der Darstellung und Entwicklung der Strukturierung des Ablaufs der Anwendung auf die sog. Flussdiagramm-Metapher zurückgreifen (z.B. Authorware von Macromedia) oder sich der Film-Metapher bedienen (z.B. Director Macromedia).

Flussdiagramm-Metapher-orientierte Autorensysteme Bei diesen Autorensystemen wird die Strukturierung, die Ablaufoptionalität, sowie die Integration unterschiedlicher Medienbausteine mittels eines graphischen Flussdiagramms mit spezifischen Knotenpunkten strukturiert. Ein Beispiel hierfür ist die Authoringsoftware Authorware 5 Attain von Macromedia, die hauptsächlich für professionelle Entwickler von Computer-based-Training sowie für Designer von Lernprogrammen konzipiert wurde: Ein Flussdiagramm verdeutlicht hierbei den Ablauf bzw. die Strukturierung einer Anwendung, ein Präsentationsfenster dient zur Darstellung und Simulation der fertigen Anwendung. Zwar dienen hier vielfältige Systemvariablen und interne Funktionen - sowie ein vorgefertigtes Set an Interaktionsstruktur-Assistenten - zur weitgehend *intuitiven* Entwicklung und Steuerung komplexer Abläufe. Auch lassen sich unterschiedliche Strukturabstraktionsstufen und verschiedene Detailrepräsentationen in der Informationshierarchie darstellen. Jedoch bleibt die für das systematische Arbeiten zwingend erforderliche Goto-Struktur bei diesem Flussdiagramm-orientierten Autorensystem weitgehend unübersichtlich. Dies hat zu Folge, dass keine eindeutige, transparente graphische Visualisierung der Verzweigungshierarchie im Sinne einer Überprüfbarkeit und Simulierung einer speziell-kontinuierlichen, zeitlichen Linie (Use Case) möglich ist.

Film-Metapher-orientierte Autorensysteme In Film-Metapher-orientierten Authoringsystemen ist es möglich auf der Grundlage zeitbasierter Abläufe weitgehend intuitiv eine Anwendung zu konzeptionieren und zu gestalten. Als Standard für professionelle Anwendungen hat sich auf dem Markt die Software Director von Macromedia etabliert. Hier ermöglicht eine Timeline die weitgehend intuitive Gestaltung von zeitbasierten Abläufen und damit die *Simulierung* einer speziell-kontinuierlichen, zeitlichen Linie (Use Case). Ein Drehbuch als zentrales Element stellt einfach verständlich dar, wann welcher Darsteller auf der Bühne erscheint. Trotz vielfältiger Möglichkeiten der Konzeption und Gestaltung interaktiver Anwendungen (Assistenzfunktionen etc.), lassen sich generell in Film-Metapher-orientierten Autorensystemen Verzweigungen und Variationen im Anwendungsablauf nicht oder nur sehr eingeschränkt graphisch visualisieren.

Den amodalen Charakter von Geschichten erfassen zum Beispiel Werkzeuge wie Plots Unlimited, Collaborator II, Story Line Pro und Dramatica. Diese Werkzeuge beschäftigen sich jedoch mit linearen Geschichten und beziehen Interaktionen des Publikums nicht in die Beschreibung der Geschichte mit ein. Das mächtigste Werkzeug in dieser Reihe ist Dramatica [Dra], Dramatica ist in Kapitel 3.1.1 beschrieben.

Zusammenfassend ist festzustellen, dass es gegenwärtig auf dem Markt zwar zahlreiche Authoringsysteme gibt, die jeweils unterschiedliche Zielsetzungen in puncto Zielgruppe, Bedienmetaphorik und insbesondere Ausrichtung der Anwendung verfolgen. Die Anforderungen, die beim Erstellen von nichtlinearen Präsentationen an ein Autorensystem gestellt werden, sind jedoch mit keinem der auf dem Markt befindlichen Autorensysteme hinreichend abgedeckt. Zwar finden sich in den jeweiligen bestehenden Autorensystemen einzelne Aspekte, die weiter verfolgt werden könnten (z.B. graphische Repräsentation des Ablaufs des Szenarios mittels einer Flussdiagramm-Metapher (Authorware, Drehbuch mit allen auftretenden Darstellern (Director), Assistenten und vorgefertigte Templates zur einfachen Generierung interaktiver Abläufe der Anwendung). Jedoch fehlt in allen Autorensystemen die Möglichkeit einer Einbindung eines zentralen Assistenzsystems, das den Autor bei der Entwicklung und Gestaltung interaktiver Szenarios unterstützt, indem es zum Einen strukturierte Informationen aus bestehenden Datenbanken oder Story Engines bereitstellt, zum Anderen die bereitgestellten Daten (z.B. auftretende Akteure, Epoche etc.) mit notwendigen Rahmenbedingungen abgleicht, konsistent in das Szenario integriert, überprüfbar macht und an eine Story Engine transferiert. Bezüglich der interaktiven Gestaltung von nichtlinearen Videopräsentationen kann schließlich keines der bekannten Autorenwerkzeuge überzeugen.

4.5 Zusammenfassung

Dieses Kapitel stellt die Grundlagen der Interaktion auf kontinuierlichen Medien vor. So werden sowohl konversationale als auch direkt manipulative Konzepte diskutiert und bewertet. Aus den Erkenntnissen dieses Kapitels wird die Schlussfolgerung gezogen, dass die Interaktionsmöglichkeiten für kontinuierliche Medien, speziell für nichtlineare Videopräsentationen, immer noch in einem frühen Stadium des Experimentierens sind. Weder sind zeitorientierte Autorenumgebungen noch konversationale Benutzerschnittstellen auf einem für den Autoren oder das Publikum zufriedenstellenden Stand. Lediglich für den Bereich der Interaktion auf Video durch Annotation wurden Ansätze entwickelt, die dem kontinuierlichen Aspekt des Mediums Rechnung tragen.

Die einzelnen Lösungen im Bereich VR-Interaktion oder Videointeraktion sind Inseln, die integriert werden sollten, um dem Benutzer einen medienzentrierten Zugang zur nichtlinearen Präsentation und damit zur nichtlinearen Narration von Geschichten zu bieten. So bieten die unter Kapitel 4.1 beschriebenen Ansätze entweder keine adequate Lösung der nichtsprachlichen Kommunikation mittels menschähnlicher Mimik und Gestik oder sie beschränken sich auf spezielle Kontexte ohne die Möglichkeit einer kontextgetriebenen Konversation zwischen System und Benutzer. Die videobasierten Ansätze in Kapitel 4.3.2 bzw. die akustischen Ansätze in Kapitel 4.2 bieten ebenfalls keine integrierten Lösungen, welche

Direkte Manipulation mit Konversation zum Nutzen des Publikums kombiniert.

Die in Kapitel 5 dieser Arbeit beschriebenen Ansätze im Bereich Konversationale Schnittstellen und Direkte Manipulation für Video bieten einen solchen integrativen Ansatz der Interaktion. Sie werden in den folgenden Kapiteln weiter ausformuliert und ihr Einsatz in verschiedenen Szenarios getestet.

5. INTERAKTIVE NARRATION VON KONTINUIERLICHEN MEDIEN

In diesem Kapitel ist die Konzeption für ein nichtlineares, kombiniert generatives und selektives Storytellingsystem, basierend auf dem Medium Video, dargestellt. Die Benutzerinteraktion ist dabei explizit als kombinierte direkt manipulative und konversationale Schnittstelle angelegt. Das Video dient als Kontext und Synchronisationseinheit der Konversation mit dem Benutzer. Das tatsächlich präsentierte Video wird von einer Story Engine zusammengestellt, die unter Berücksichtigung von Dramaturgie, Spannung und Immersion entsprechende Storyelemente - in diesem Falle Video-Clips - interaktiv beeinflusst durch den Benutzer, zusammenstellt. An die Story Engine ist ein Autorensystem angeschlossen, das die Erstellung einer nichtlinearen Dramaturgie für einen Autoren erleichtert. Die - dem Interaktionskonzept auf die Story zugrunde liegende - Conversation Engine beinhaltet ein Conversation Modell, das automatisch u.a. auf Turntaking, Historie und Kontext der Konversation des Systems mit dem Benutzer eingeht.

Das Konzept wird in die folgenden Module gegliedert:

Nichtlineare Videopräsentation Die Grundlagen der Präsentation eines nichtlinearen Videos werden erläutert. Hierbei wird unter anderem auf die Client-Server Struktur eingegangen sowie die dazu passende Middleware diskutiert.

Benutzerinteraktion Die Benutzerinteraktion sowie deren Ästhetik werden diskutiert. Dabei wird ein Konzept für die Interaktion mit kontinuierlichen Medien aufgezeigt, dass in einen direkt manipulativen und einen konversationalen Teil gegliedert ist.

Konversation Die konversationale Modellierung wird von einer speziellen Conversation Engine übernommen, die das konversationale Verhalten der Maschine in Relation zu Content, Kontext und Benutzerinteraktion steuert. Dies geschieht auf einer expliziten symbolischen Ebene und wird amodal modelliert. Neben dieser Ebene wird die entsprechende Präsentationsschicht sowie die Interpretation der Benutzereingaben angerissen, die eine modale Modellierung repräsentiert.

Direct Manipulation Die direkten Aktionen des Benutzers auf den präsentierten Medien werden medienspezifisch durch die jeweilige Applikation verwirklicht. Im Falle des Mediums Video wird eine explizite, temporale und intramediale Annotation des Videos realisiert.

Nichtlineare Erzählstruktur Die zugrunde liegende Story Engine wird aufgespannt und diskutiert. Hierbei wird auf die Zusammenstellung der generischen und selektiven Kom-

ponenten der Präsentation eingegangen. Der Ansatz von Propp legt die Basis für die nichtlineare, interaktive Präsentation von Geschichten.

Anschließend werden die Implementierungsgrundlagen diskutiert und das Gesamtsystem in seiner Architektur beschrieben.

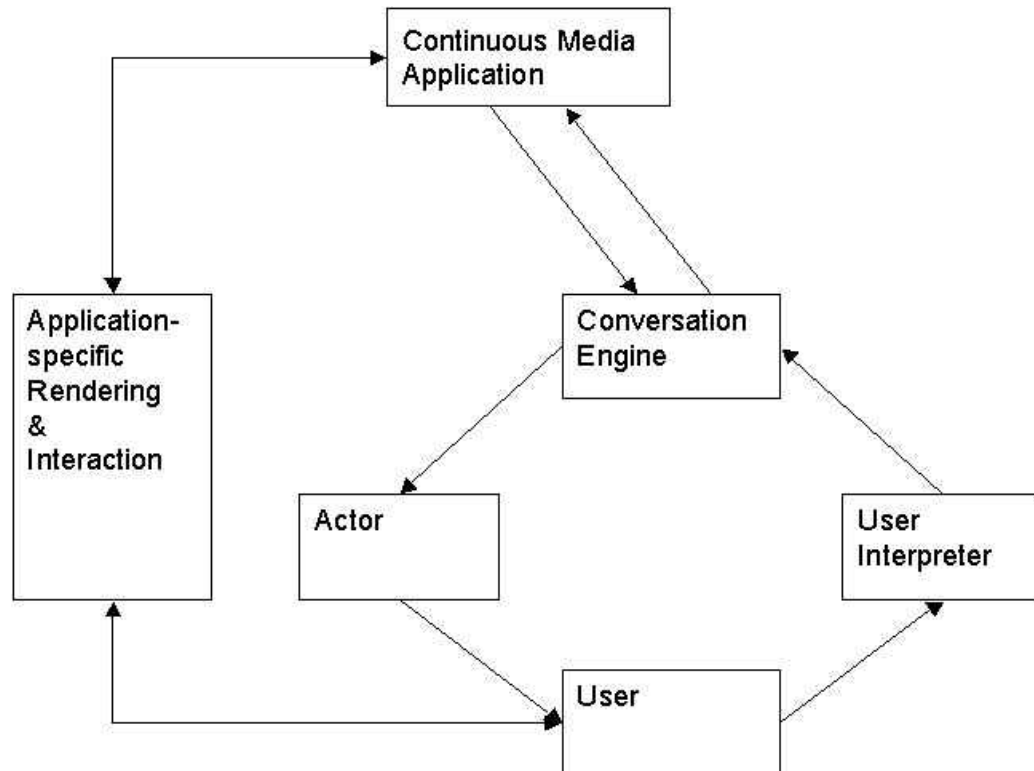


Fig. 5.1: Konzept Trennung applikationsspezifisches Rendering & direkt manipulative Interaktion in Kombination mit konversationaler Interaktion

Das Konzept sieht eine Trennung zwischen inhaltsbezogenen Komponenten und User Interface bezogenen Komponenten vor. Die Story-generierung und die Content-Datenhaltung wird als Zulieferer der konversationalen Interaktion, repräsentiert durch ein konversationales Modell innerhalb einer Conversation Engine (CE), gesehen, vgl. [Bra02b]. Die Darstellung der Inhalte und konversationaler Gesprächspartner wird als vorgeschaltetes Element der CE angesehen, ebenso die Interpretation der Benutzereingabe. Das Konzept ist im Bild 5.1 dargestellt.

Neben den rein konzeptbezogenen Ausführungen wird in diesem Kapitel die Grundlage eines API für nichtlineares Geschichtenerzählen¹ gelegt. Dieses API stellt für die verschiedenen Bereiche des nonlinear Storytelling - Konversation, Narration, medienspezifische Interaktion

¹ Eine solche API wird erstmals von Encarnação gefordert, vgl. [ES99, Enc97].

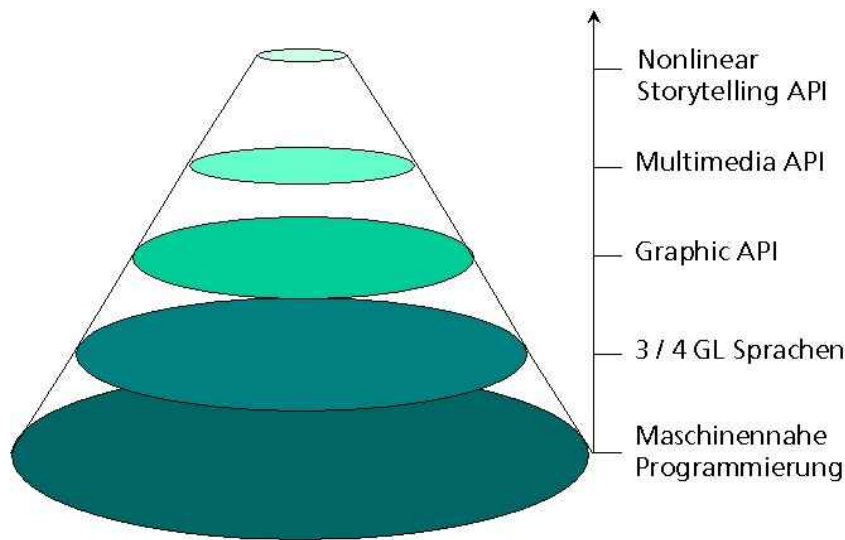


Fig. 5.2: Konzept API für nichtlineares Storytelling - Anwendungsunterstützung durch Fokussierung der Möglichkeiten

- XML²-basierte Schnittstellen zur Verfügung³. Das API zeichnet sich durch eine hohe Stufe der Abstraktion aus. Dies führt zu einer Fokussierung der Möglichkeiten des API-Benutzers (Anwendungsprogrammierer). Die Fokussierung führt dazu, dass der Anwendungsprogrammierer nicht selbst zur Lösung der Problematiken von nichtlinearem Storytelling gezwungen wird - dies wird von den durch das API zugreifbaren Komponenten des nichtlinearen Narrationssystems für ihn übernommen. Die Fokussierung des API bringt dem Anwendungsprogrammierer eine *vernünftige, funktionierende*⁴ Ästhetik für seine narrativen Anwendungen, vgl. Braun [Bra02e]. Das API ist von Braun in [Bra01c] definiert.

5.1 Applikationsspezifisches Konzept Video

Der nichtlineare Videoservice (die Basis der Präsentation von Video-Clips) ist als externe Applikation realisiert. Notwendigerweise erfüllt der Videoservice die Anforderungen, die im Kapitel 3.4 an ihn gestellt sind. Diese werden zur Erinnerung hier rezitiert:

- Zumindest Branching-basierter Aufbau der Videopräsentation.
- Zumindest halbautomatische Zusammenstellung von Video-Clips.
- Intramediale, explizite Interaktionsmöglichkeit auf graphische und akustische Videoinhalte.

² Vgl. [BPSMM00].

³ Diese werden im Kapitel 6 zur Validierung des Konzeptes mittels entsprechender Applikationen genutzt.

⁴ Die Bewertungen Vernünftig und Funktionieren sind aus dem Fach Design abgeleitet. So funktioniert eine Anwendung, wenn der durch den Autor angestrebte Effekt beim Publikum eintritt.

Wird Video als kontinuierliches Medium gewählt, so ist der Video-Clip die kleinste Kontexteinheit des Systems. Er wird mittels Metadaten bez. der synchronen hypermedialen Annotation sowie mit allgemeinen Informationen zum Inhalt des Video-Clips versehen. Dies erlaubt sowohl eine Anbindung an das konversationale Teilsystem als auch eine Anbindung an das Story Engine Teilsystem.

Die Metainformationen eines jeden verwendeten Video-Clips, z.B. Inhaltsinformationen, werden in einem applikations-spezifischen Format⁵ abgelegt, ebenso die graphischen und akustischen Annotation der Video-Clips⁶.

5.2 Benutzerinteraktion

Die Benutzerinteraktion spaltet sich in einen direkt manipulativen und in einen konversationalen Teil. Tatsächlich werden jedoch die direkt manipulativen Interaktionen des Benutzers ebenfalls in der Conversation Engine verarbeitet, da auch diese Eingaben einen Einfluss auf die konversationale Modellierung besitzen.

5.2.1 (Video-)Applikationsspezifischer Direkt Manipulativer Zugang

Der direkt manipulative Zugriff auf Medien ist in extremer Weise abhängig von den Medien selbst. Da im diskutierten System eine videobasierte Darstellung angestrebt wird, werden videospezifische Interaktionsmöglichkeiten⁷ angeboten, für die eine intuitive, den zeitlichen Aspekten des Mediums Video gerechte, synchrone, direkt manipulative Interaktionsmöglichkeit⁸ in diesem Kapitel definiert wird. Um die künstlerische Ästhetik des Videos und die Ausdrucksvielfältigkeit des Autoren nur minimal zu beschneiden wird der Ansatz der intramedialen und expliziten Annotation von graphischen und akustischen Teilen des Videos als Konzept der synchronen Interaktion gewählt.

Die Annotation des Videos entspricht dabei den Anforderungen, die in Kapitel 4.3.2.1 aufgestellt sind:

- Die Annotation muss beim Start die gesamte verbleibende Interaktionszeit anzeigen.
- Die Annotation muss während der Dauer der Annotation die verbleibende Interaktionszeit anzeigen.
- Die Annotation muss zum Ende der Annotationszeit explizit das Ende der Annotation anzeigen.

⁵ Zum Beispiel in dem Format UIML nach der Spezifikation von Blechschmitt, vgl. [BB01]. Inhalte, welche anhand dieser Spezifikation generiert sind, erweisen sich als direkt interpretierbar durch einen Dialoginterpret, vgl. [BS01a], [Ble01], welcher mit der konversationalen Modellierung, siehe Kapitel 5.2.2, kommuniziert.

⁶ Siehe Kapitel 6.3.2.

⁷ Die Grundlagen dieser Interaktionsmöglichkeiten sind in Kapitel 4.3.2 diskutiert.

⁸ Vgl. Kapitel 4.3.3

Weiterhin erfolgt die Annotation gemäß den Anforderungen in Kapitel 4.3.2.1 intramedial und explizit:

- Wird die graphische Ausgabe eines Videos annotiert, so erfolgt die Annotation durch eine graphische Markierung des Informationsobjektes im Video. Die für den Benutzer logische Interaktion ist ein Mausklick (als Repräsentanz einer Zeigegeste) auf die Annotation und damit auf das Objekt.
- Wird die akustische Ausgabe eines Videos annotiert, so erfolgt die Annotation durch eine akustische Markierung des Informationsobjektes im Video. Die für den Benutzer logische Interaktion ist eine akustische Antwort auf die Annotation. Diese akustische Antwort erfolgt für den Benutzer quasi im gleichen Medium, d.h. im akustischen Kanal.

Neben dieser synchronen Interaktionsmöglichkeiten soll der Benutzer die Möglichkeit besitzen, die von Video gewohnte Möglichkeit der sogenannten *Videorecorder Funktionalität* als asynchrone, direkt manipulative Interaktionsmöglichkeit auf das Video zu nutzen.

Aus Benutzersicht werden minimal die Funktionalitäten in Tabelle 5.1 gefordert, die entweder durch die synchrone oder asynchrone Interaktionsmöglichkeit des Benutzers mit dem Video verwirklicht werden sollen. Die Aktion, die durch das Auswählen eines Hyperlinks aktiviert wird, ist völlig frei wählbar. Die Art der Aktion wird unterschieden in drei Kategorien:

- Videobezogen: Die Videopräsentation selbst ändert sich, d.h. eine navigatorische Interaktion auf das Video, welche eine Änderung der aktuellen Videopräsentation bewirkt. Die Videorecorder-Funktionalität wird dabei nicht als Änderung der aktuellen Präsentation angesehen.
- Conversationsbezogen: Die Interaktion auf das Video erzeugt eine konversationale Ausgabe des Systems.
- Sonstige: Die Interaktion auf das Video erzeugt eine sonstige Ausgabe durch das System - beispielsweise werden zusätzliche mediale Informationen angezeigt.

Eine synchrone oder asynchrone direkt manipulative Interaktion wird zwingend an den konversationalen Teil des Systems als Eingabe des Benutzers gemeldet⁹.

Die Videorecorderfunktionalität ist im Wesentlichen in jedem Videopräsentationswerkzeug enthalten und somit anwendungsspezifisch gelöst. Die Annotationsinformation ist ebenfalls Anwendungsspezifisch abgelegt¹⁰, jedoch im Allgemeinen nicht explizit durch die Anwendung gegeben. Entsprechende File-Format-Plugin und Rendering-Plugin müssen applikationsspezifisch¹¹ entwickelt werden.

⁹ Vgl. Kapitel 5.2.2.

¹⁰ Vgl. Anhang A.2.

¹¹ Vgl. Kapitel 6.3.2.1.

Tab. 5.1: Interaktive Funktionalitätsanforderungen für Video

Funktion	Beschreibung
Wähle Video	Auswählen einer Videopräsentation.
Start	Das gewählte Video wird gestartet.
Stop	Das ablaufende Video wird gestoppt.
Pause	Das ablaufende Video wird angehalten und kann durch erneutes Anwählen von Pause und Start weiter angesehen werden.
Vorspulen	Erlaubt dem Anwender das Vorspulen des Videos.
Rückspulen	Erlaubt dem Anwender das Zurückspulen des Videos.
Hyperlink aktivieren	Aktivieren eines Hyperlinks im Video. Dies kann auf zumindest zwei Arten erfolgen: durch die sprachliche Ansteuerung des akustischen Hyperlinks oder durch Anwählen eines graphischen Hyperlinks mit der Maus.

5.2.2 Konversationaler Zugang

Konversationen bestehen aus diversen Faktoren, die sich direkt aus der zwischenmenschlichen Kommunikation - daher der Mensch-Mensch Kommunikation - ableiten. Diese Faktoren sind in einer ersten Stufe amodal¹², d.h. ohne einen Bezug zur medialen Ausprägung der Konversation, modelliert. Im Anschluss wird eine modale Modellierung für einen konversationalen Charakter (in Form eines Avatars) vorgestellt¹³, vgl. Braun [Bra02f].

5.2.2.1 Symbolische Modellierung

Aus den Anforderungen an die konversationalen Aspekte werden folgende Basisaspekte, welche zur Bildung eines *Konversationsmodells* herangezogen werden, vgl. Braun [BS01b], abgeleitet:

- Konversationsteilnehmer
 - name: Eindeutiger Name des Konversationsteilnehmers.
 - typ $\in \{\text{aktiv, passiv}\}$: Zeichnet den Konversationsteilnehmer als eher aktiven oder eher passiven Gesprächspartner aus.
 - rederecht $\in \{\text{true, false}\}$: Gibt an, ob der Konversationsteilnehmer in der aktuellen Situation das Rederecht besitzt.

Dieser Aspekt beschreibt einen Konversationsteilnehmer. Ein Beispiel hierfür ist etwa der Benutzer selbst oder ein virtueller Actor, eine virtuelle Persönlichkeit bzw. Virtual Human. Die mediale Charakteristik des Konversationsteilnehmers findet keine Berücksichtigung. Jeder Konversationsteilnehmer wird eindeutig über seinen Namen identifiziert.

¹² Vgl. Kapitel 5.2.2.1.

¹³ Vgl. Kapitel 5.2.2.2

- Konversationsverhalten

- Verhalten $\in \{ \text{EroeffnungDiskurs, WechselnDiskurs, RederechtAnfordern, RederechtHolen, Reden, RederechtAnbieten, RederechtAbgeben, Zuhören, StartSequenz, WechselSequenz, EndSequenz, AbschließenDiskurs} \}$: Hier wird das konversationale Verhalten eines Konversationsteilnehmers spezifiziert.
- sender: Der Name des Konversationsteilnehmers.
- empfaenger-liste: Eine Liste von Konversationsteilnehmern, an die das Verhalten gesendet wird.
- inhalt: Ein Verweis auf den Inhalt, auf den sich das Verhalten bezieht.
- diskurs: Ein Verweis auf den Diskurs, innerhalb dessen sich das Verhalten abspielt.
- zeitpunkt: Der Beat, zu dem das Verhalten auftritt (siehe Aspekt Time).
- intensitaet $\in \{ \text{forcieren, neutral, daempfen} \}$: Die Intensität des Verhaltens.

Dieser Aspekt bedeutet vereinfacht ausgedrückt: Jemand (der Sender) kommuniziert etwas (das Verhalten und den Content) zu einem Anderen (die Liste der Adressaten) innerhalb eines Diskurses zu einem bestimmten Zeitpunkt.

- Inhalt

- name: Eindeutiger Bezeichner des Contents.
- typ $\in \{ \text{Frage, Antwort, Ausdruck} \}$: Der Typ des Inhalts spezifiziert den lokutionären¹⁴ Akt des Inhalts.
- referenz: Ein Bezug des Inhalts.
- ziel: Der Konversationsteilnehmer, an welchen der Inhalt gerichtet ist.
- status $\in \{ \text{NichtBegonnen, Bearbeitung, Unterbrochen, Fertig} \}$: Der Bearbeitungsstatus des Inhalts.
- diskurs: Verweis auf den Diskurs, innerhalb dessen der Inhalt bearbeitet werden soll.
- dringlichkeit $\in \{ \text{wenig, mittel, viel} \}$: Die Dringlichkeit des Inhalts.
- wichtigkeit $\in \{ \text{wenig, mittel, viel} \}$: Die Wichtigkeit des Inhalts.
- emotionalitaet $\in \{ \text{gut, mittel, schlecht} \}$: Die emotionale Bewertung des Inhalts.

Der Inhalt spezifiziert einen illukutionären¹⁵ Akt. Er kann spezialisiert werden zur Antwort (d.h. mit einer Beziehung zu einem früheren Konversationselement), zur Frage (mit einer Beziehung zu einem später noch folgenden Konversationselement), ohne Spezialisierung ist der Inhalt ein Ausdruck (im Bezug zur momentanen Situation der Konversation).

¹⁴ Vgl. Kapitel 4.1.

¹⁵ Vgl. Kapitel 4.1.

- Antwort erweitert Inhalt

- `wiederholungen_erlaubt` $\in \{\text{true}, \text{false}\}$: Die Eigenschaft gibt an, ob eine Antwort mehrfach gegeben werden darf.

Ein Inhalt kann als eine Antwort spezialisiert sein. Sein Typ muss jedoch nicht zwangsläufig eine Antwort sein. Diese Erweiterung von Inhalt gibt die tatsächliche Interpretation des Inhaltselements an - eine Antwort kann z.B. den Typ *Ausdruck* besitzen, dient aber z.B. als Informationsweitergabe an eine angeschlossene Applikation.

- Frage erweitert Inhalt

- `wiederholungen`: Die Eigenschaft *wiederholungen* zeigt an, wie oft eine Frage schon wiederholt wurde.

Ein Inhalt kann als Frage spezialisiert werden. Sein Typ muss jedoch nicht zwangsläufig eine Frage sein. Diese Erweiterung von Inhalt gibt die tatsächliche Interpretation des Inhaltselements an - eine Frage kann z.B. den Typ *Ausdruck* besitzen, dient aber als Auslöser eines Antwortverhaltens.

- Story

- `name`: Eindeutiger Bezeichner der Story.
- `typ` $\in \{\text{sequence}, \text{asynchronous}\}$: Charakteristik der Story.
- `inhalts-liste`: Liste von Namen von Inhalts-Elementen oder von Spezialisierungen von Inhalts-Elementen
- `status` $\in \{\text{Nichtgestartet}, \text{Gestoppt}, \text{Name}\}$: Der Status gibt an, ob eine Story schon begonnen wurde. In diesem Fall wird der Name des entsprechenden Inhalts-Elements in *status* verzeichnet.

Eine Story wird als übergeordnete Struktur für Inhalts-Aspekte angesehen. Für die Konversation ist die Art der Beziehung nicht wichtig - genausowenig ist es wichtig, was der Inhalt wirklich beschreibt. Interessant aus Sicht der Konversationsmodellierung ist lediglich ob die Struktur eine lineare oder nichtlineare Charakteristik besitzt - d.h.: wird eine Sequenz von Informationen beschrieben oder stehen die Informationen in einem asynchronen Zusammenhang?

- Denkprozess

- `referenz`: Der Name eines Inhalts-Elements.

Dies ist eine Information über einen Konflikt, welcher in der konversationalen Modellierung herrscht. Der Grund des Konflikts wird in der Referenz angegeben.

- Time

- `zeitpunkt-liste`: Liste aller Beats.

- aktueller_zeitpunkt: Aktueller Beat.
- naechster_zeitpunkt: Nächster Beat.

Die zeitliche Ausdehnung einer Konversation wird über Beats¹⁶ abgewickelt. In der Eigenschaft aktueller_zeitpunkt steht dabei der letzte Zeitpunkt, zu dem ein Verhalten aufgetreten ist. In naechster_zeitpunkt steht der nächste mögliche Zeitpunkt, zu dem ein Verhalten auftreten kann. Die Gesamtheit aller bisherigen Zeitpunkte wird in der Eigenschaft zeitpunkt-liste gesammelt.

- Diskurs

- name: Eindeutiger Bezeichner des Diskurses.
- zeitpunkt: Start-Beat des Diskurses.

Verschiedene Diskurse können von ein und derselben Conversation Engine gehalten werden. Jeder Diskurs wird dabei durch seinen Namen beschrieben und bekommt einen Startzeitpunkt zugewiesen.

Mit diesen - relativ simplen - konversationalen Aspekten können sehr komplexe konversationale Situationen dargestellt werden. Die Darstellung des Verhaltens von einzelnen Konversationsteilnehmern geschieht über das Einfügen von entsprechenden Aspekten in die konversationale Situation. Dies entspricht einer Abbildung von konversationalen Ausgangssituationen in entsprechende konversationale Endsituationen (die Endsituationen repräsentieren die durch das Verhalten bewirkte Veränderung). Die Generierung von Aspekten für einen künstlichen Konversationsteilnehmer wie zum Beispiel einen Virtual Human oder Synthetic Actor wird durch *Regeln* wahrgenommen. Diese Veränderung der konversationalen Situation durch Hinzufügen, Ändern oder Entfernen von konversationalen Aspekten spiegelt dabei das Verhalten des künstlichen Konversationsteilnehmers wider. Jeder Konversationsteilnehmer besitzt seine eigene Menge an Regeln, die spezifisch auf seine konversationale Situation eingehen.

Eine Regel besitzt folgenden grundsätzlichen Aufbau:

- LHS (Left Hand Side): Die linke Seite der Regel beschreibt eine konversationale Teilsituation. Sie muss daher nicht alle Aspekte einer konversationalen Situation berücksichtigen.
- \implies : Der Pfeil beschreibt eine Implikation, welche sich aus der LHS ergibt.
- RHS (Right Hand Side): Die rechte Seite der Regel beschreibt das Hinzufügen, Modifizieren oder Entfernen von Aspekten aus der konversationalen Situation, welche die LHS als konversationale Teilsituation besitzt.

¹⁶ Ein Beat ist die kleinste Einheit, innerhalb derer ein für die Konversation relevanter Aspekt auftreten kann.

Die Regelmengende ist - trotz der überschaubaren Anzahl der Aspekte - recht groß, jedoch einfach zu modifizieren und zu erweitern: Da Regeln ausschließlich konversationale Teilsituationen beschreiben (konversationale Situationen *matchen*¹⁷) und keine prozedurale Abfolge vorgeben, kann die Regelmengende durch einfaches Hinzufügen von Regeln erweitert werden. Eine Regel selbst kann durch simples Hinzufügen von Aspekten auf der LHS bzw. der RHS erweitert werden, ohne dass sonstige Aspekte der LHS bzw. RHS verändert werden müssen.

Im folgenden ist eine Regel beispielhaft notiert. Die Notation der Regel geschieht dabei angepasst an den Zweck der Regel - dieser ist das Matchen von Aspekten:

- Jeder Aspekt, welcher in der LHS gematcht werden soll, wird mit seinem Namen bezeichnet. Seine Eigenschaften werden im Anschluss angegeben mit dem Wert, den die Eigenschaften besitzen sollen.
- Die Eigenschaften der Aspekte müssen nicht vollständig aufgelistet sein - eine Eigenschaft eines Aspektes, welche keine Berücksichtigung findet, darf jegliche Ausprägung besitzen.
- Aspekte, welche in der RHS verändert werden sollen, müssen in der LHS gemacht werden und bedürfen einer Referenzierung in der LHS, ausgedrückt durch einen \Leftarrow .
- Prädikate, Funktionen etc. sind vereinfacht und umgangssprachlich ausgedrückt.
- Variablen sind mittels eines ? markiert - sie werden zum Zeitpunkt Ihres ersten Auftretens gebunden.
- Konstanten sind in Großbuchstaben notiert.
- Kommentare sind mit einem ; markiert.
- Die Regel wird von oben nach unten ausgewertet.

Die beispielhafte Regel beschreibt eine Situation, in der ein spezialisierter Inhalt - eine Antwort - mit einem Bezug zu einem weiteren spezialisierten Inhalt - einer Frage - existiert und unter Berücksichtigung einiger Rahmenbedingungen durch den Konversationsteilnehmer zum Ausdruck gebracht werden soll.

(Rule Antwort_auf_Frage

```

; Der erste Aspekt, welcher in der Situation matchen soll,
; ist das Rederecht des Konversationsteilnehmers. ?*ICH*
; beschreibt dabei eine Variable, in der der Name des
; Konversationsteilnehmers steht, welcher durch die Regelmengende
; repräsentiert wird.
(Konversationsteilnehmer
```

¹⁷ Das Wort match wird im folgenden eingedeutscht benutzt.

```

        (name ?*ICH*)
        (rederecht TRUE)
    )

; Hier wird eine Antwort gesucht, welche sich
; auf die Frage ?Frage_des_Benutzers bezieht.
?Aspekt_Antwort<-
    (Antwort
        (name ?name_Antwort)
        (referenz ?Frage_des_Benutzers)
        (diskurs ?Diskurs_Antwort)
        (status NICHT_BEGONNEN)
    )

; Das ist der Inhalt, welcher durch ?Frage_des_Benutzers
; uebermittelt wird. Der Inhalt sollte eine Frage sein -
; der Typ des Inhalts jedoch ist unrelevant.
(Frage
    (name ?Frage_des_Benutzers)
    (discourse ?Diskurs_Antwort)
)

; Im zur Frage zugeordneten Verhalten steht, wer die Frage gestellt hat
; An diesen muss die Antwort gegeben werden.
(Konversationsverhalten
    (sender (?Empfaenger_Antwort, nicht ?*ICH*))
    (inhalt ?Frage_des_Benutzers)
    (empfaenger-liste (?EmpfaengerListe_der_Frage enthält ?*ICH? ))
    (diskurs ?Diskurs_Antwort)
)

; Feststellen, ob schon eine Begrueßung des Fragenden vorliegt.
(Konversationsverhalten
    (verhalten EROEFFNUNG)
    (sender (?Sender_Eroeffnung bin ?*ICH*))
    (empfaenger (?Empfaenger_Begrueßung enthaelt ?Empfaenger_Antwort)
    (diskurs ?Diskurs_Antwort)
)

; Die Story muss weiter gesetzt werden!
; Wenn die aktuelle Story schon bei diesem Inhalt ist,
; so muss der Inhalt nicht mehr beantwortet werden.
?Aspekt_Story_Antwort<-
    (Story

```

```

        (inhalt (?Inhalt_Story ist ?Name_Antwort))
        (status (?Status_Story ist nicht ?Name_Antwort))
    )

=>

; Erzeuge Verhalten, das Inhalt wiedergibt.
; Setze die Zeit eins weiter...

(Einfuegen
  (Konversationsverhalten
    (verhalten REDEN)
    (sender ?*ICH*)
    (empfaenger-liste ?Empfaenger_Antwort)
    (inhalt ?Name_Antwort)
    (diskurs ?Diskurs_Antwort)
    (zeitpunkt (Nächster Zeitpunkt))
  )
)

(Veraendern ?Aspekt_Antwort
  (status BEARBEITUNG)
)
(Veraendern ?Aspekt_Story_Antwort
  (status ?Name_Antwort)
)
)

```

Regeln dieser Art sind für konversationale Verhaltensmengen und Verwaltungsprozesse erstellt. Die verschiedenen Regeln werden zur besseren Übersicht gruppiert, die entsprechenden Gruppen aufgelistet und erläutert.

- Verarbeitung von konversationalen Aspekten
 - Storyverwaltung: Verwaltung der Inhalte, welche dem Konversationsteilnehmer zur Vermittlung vorgegeben werden.
 - Diskursverwaltung: Der Status der verschiedenen Diskurse.
 - Rederechtverwaltung: Verschiedene Konversationsteilnehmer können das Rederecht explizit anfordern oder tatsächlich holen bzw. durch Aktivitäten implizit erringen, siehe Abbildung 5.3.
 - Kontextverwaltung: Aufmerksamkeit, Hintergrundlautstärke, Benutzervorgaben, etc..
- Generierung von konversationalen Aspekten

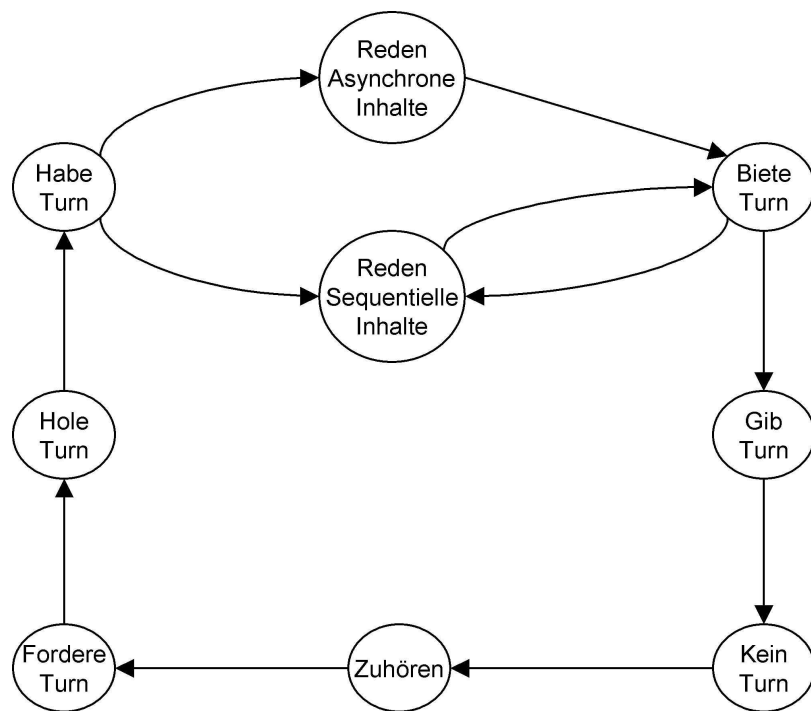


Fig. 5.3: Skizze des Übergangsmodells der konversationalen Rederecht-Verwaltung

- Turntaking (Rederecht-Verhalten): Rederecht wird angefordert, zur Verfügung gestellt, abgegeben, geholt.
- Diskursverhalten: Diskurse werden eröffnet, abgeschlossen, gewechselt, wieder aufgenommen.
- Asynchrones Verhalten: Fragen, Antworten, Unterbrechungen.
- Synchrones Verhalten: Bearbeiten von zusammenhängenden Inhalten.

Die Dringlichkeit und die Wichtigkeit der Inhalte entscheiden dabei mit darüber, wann ein Inhalt ausgegeben wird. Generell werden die Informationen, welche dringend sind, vor anderen Informationen ausgegeben. Informationen, welche wichtig sind, werden ebenfalls vor anderen Situationen ausgegeben. Die beiden Faktoren beeinflussen sich auch gegenseitig: Dringende, wenig wichtige Informationen zum Beispiel werden anders ausgegeben als wenig dringende, dafür aber sehr wichtige Informationen.

Die Emotionalität einer Nachricht, d.h. die Bewertung, ob ein Inhalt als positiv oder negativ zu werten ist, beeinflusst ebenfalls die Art der Regelauswertung - schlechte Nachrichten werden zum Beispiel in der zwischenmenschlichen Konversation mit einer verlängerten Einleitung übertragen.

Der Auswertungsprozess der Regeln auf einer gegebenen konversationalen Situation geschieht in der Regel durch ein Inferenzsystem¹⁸, wie es aus der Künstlichen Intelligenz (KI) bekannt ist, vgl. [PMG98].

5.2.2.2 Charakter

Die amodalen Befehle der Conversation Engine müssen in eine entsprechende modale Ausformulierung transferiert werden. Hierzu dient eine Charakter Engine, die einen entsprechenden Charakter ansteuert. Die Charakter Engine zeichnet sich durch eine hohe Flexibilität in Hinsicht des darzustellenden Avatars - sowohl extrem menschähnliche als auch comichaft-anthropomorphe Darstellung des Avatars - aus.

Das Avatar System erkennt Befehle auf verschiedenen Abstraktionsstufen und kann sowohl die verschiedenen Abstraktionsstufen als auch sprachsynchrone Lippenbewegungen mischen. Vier verschiedene Abstraktionslevel werden unterschieden, vgl. Braun [BARM00]:

- Motivation: Kommandos, die ein pro-aktives Verhalten auslösen wie *redselig, überaktiv, traurig, wütend*
- Aufgaben (Task): Kommandos, die direkte Aktionen auslösen wie *Begrüßung, Anzeigen, GiveTurn*. Auf dieser Ebene findet die modale Umsetzung der amodalen Conversation-Engine-Befehle statt.
- Eigenschaften (Feature): Diese Kommandos lösen eine Modifikation von abstrakten Körper-Komponenten des Avatars aus - wie zum Beispiel *linke Augenbraue hoch* oder *Mund lächeln*.
- Geometrie: Diese Kommandos lösen eine direkte Modifikation an den Polygonen des Avatar-Modells aus.

Mittels dieses Ansatzes lassen sich die Verhaltensbefehle des Avatars sehr einfach durch einen Autoren zusammensetzen sowie sehr einfach zu einer Anwendung übertragen. Tatsächlich aber wird aufgrund dieses Ansatzes ein generisches System implementiert: Animationssequenzen des Avatars müssen nur ein einziges Mal definiert werden und können dann in beliebiger Weise miteinander kombiniert werden. Zusätzlich kann die komplette Geometrie des Avatars ausgetauscht werden ohne die Animationssequenzen oder Animationsbefehlssequenzen modifizieren zu müssen.

Die abstrakten Verhaltensbeschreibungen für den Avatar sind in Form einer XML-Struktur gespeichert notiert.

Beispiel XML-Struktur:

```
<avatar name "default">
  <motivation name="friendly" intensity="0.7" start="0" duration="0">
```

¹⁸ Die entsprechenden Systeme werden im Kapitel 6 beschrieben.

```

    <task characteristic "wait" target "user">
  </task>
</motivation>
</avatar>

```

Diese Beschreibung veranlasst den Avatar dazu, freundlich auf die nächste Benutzereingabe zu warten. Startzeit und Dauer des Motivation-Befehls sind hier auf Null gesetzt, da das Verhalten bis zur Benutzeraktion beibehalten werden soll. Detailverhalten wie ständige Kopfbewegung und unwillkürliches Blinzeln wird im Verhaltensübersetzer automatisch erzeugt und mit dem definierten Verhalten gemischt. Für die Informationen auf Motivations- und Feature-Ebene kann vom Autor auch der Intensitätsgrad angegeben werden. Für die Task-Ebene ist diese Vorgangsweise nicht sinnvoll, da man beispielsweise nicht halb oder völlig präsentieren kann. Andererseits ist die Möglichkeit einer Abstufung zwischen 0 % *glücklich* und 100 % *glücklich* für den Autoren wünschenswert.

Die konversationalen Aspekte der Conversation Engine werden durch verschiedene Animationen¹⁹ umgesetzt, die sich aus den Studien der Kameraperspektiven²⁰ und der Grundüberlegungen²¹ zur Konversation ergeben.

5.3 Nichtlineare Erzählstruktur

Videopräsentationen besitzen ursprünglich einen rein linearen Charakter der Narration. Der Charakter der Narration wird durch die Verwendung von Video-Clips, die durch Autorenvorbestimmte Publikumsinteraktion zu einer gewissen poly-Linearität erweitert. Um eine vollständige nichtlineare Erzählstruktur zu erreichen ist eben diese als Modell zu notieren und durch eine Story Engine auszuführen.

Mittels des Geschichtenmodells von Propp ist eine sehr flexibel einsetzbare Grundlage zur nichtlinearen Storynarration entwickelt worden, siehe Abbildung 5.4, vgl. Braun [Bra02d]. Diese Grundlage dient einer ausführenden Einheit, der Story Engine, als Basis ihrer Entscheidungen. Das generische Konzept einer Story Engine wird von Propp selbst angeregt²²:

„It is possible to artificially create new plots of an unlimited number. All of these plots will reflect a basic scheme, while they themselves may not resemble one another. In order to create a tale artificially, one may take and A, then one of the possible B's then a C↑, followed by absolutely and D, then an E, then one of the possible F's, then any G, and so on. In doing this, any elements may

¹⁹ So wird beispielsweise die Technik des Turntaking durch folgendes Verhalten visualisiert: Beim Abgeben des Turn sieht der Avatar den Benutzer an, um ihm visuell eine Möglichkeit des Augenkontaktes zu bieten, der dem Benutzer die geordnete Übernahme des Gespräches ermöglicht. So wird jeder Befehl, der von der Conversation Engine an die Character Engine gerichtet wird, durch diese explizit animiert.

²⁰ Vgl. Kapitel 3.1.2.

²¹ Vgl. Kapitel 4.1.3.

²² Vgl. [Pro58], S. 111.

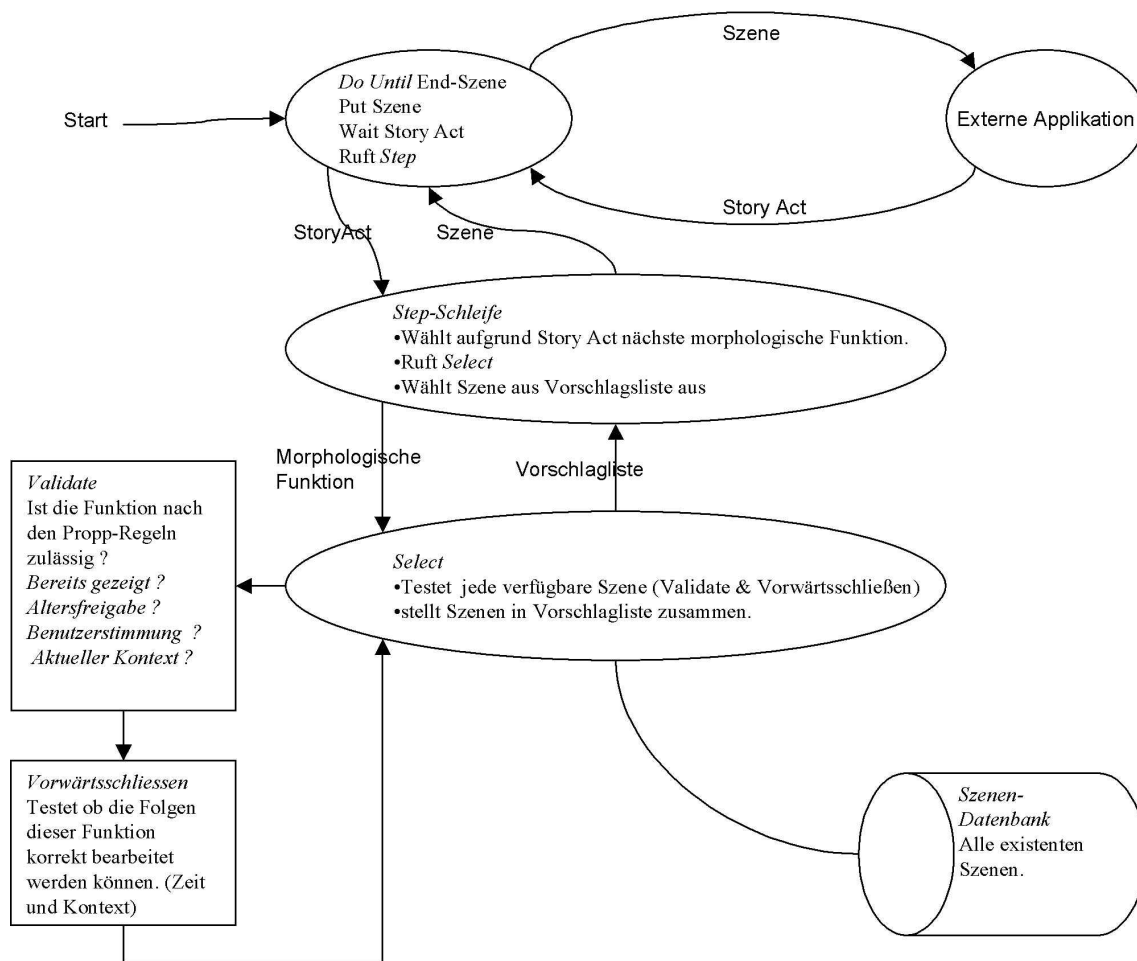


Fig. 5.4: Simplifizierte Skizze der morphologischen Story Engine

be dropped (except possibly for A or a), or repeated three times, or repeated in various forms.“

Abgeleitet von Propps Basisvorstellung, erreicht die Story Engine eine nichtlineare Narration auf zwei Abstraktionsebenen:

- Zusammenstellung von Funktions²³-Sequenzen: Auf oberster Ebene werden, in Echtzeit, Sequenzen von Funktionen zusammengestellt, die eine dramatische Handlung ergeben. Nach jeder Interaktion bzw. sonstigen Änderung von story-relevanten Parametern wird die Sequenz neu berechnet. Die interaktive Geschichte ist konsistent, solange zumindest eine mögliche Sequenz von Funktionen die Geschichte zu einem definierten Ende führt.

²³ Hierbei wird die Begrifflichkeit der Funktion adäquat zur Definition von Propp genutzt.

- Abbildung von Funktionen auf Szenen: Jede Funktion²⁴ muss auf eine Szene (z.B. einen Video-Clip, eine Animation, eine Textpassage) abgebildet werden, um letztlich zur Darstellung zu kommen.

Der Begriff der Funktion und der Szene werden im folgenden weiter vertieft.

5.3.1 Morphologische Funktionen

Das Konzept der Funktionen nach Propp sowie die entsprechende Anweisung von Propp zum Zusammenstellen von Funktionssequenzen implizieren im eigentlichen Sinne eine lineare Geschichte ohne Einflussnahme des Zuschauers/Benutzers. Um die Einflussnahme des Zuschauers zur Laufzeit der Geschichte zu gewährleisten, muss das System von Propp um weitere Konzepte erweitert werden:

- Varianten: Ein Konzept, um den geschichtlichen Ausgang einer Funktion variabel anpassbar halten zu können.
- Polymorphie: Die Möglichkeit, einer Szene mehrere Funktionsmöglichkeiten zuzuordnen um diese dann zur Laufzeit auszuwählen. Die Funktion einer Szene wird im Zusammenhang der Geschichte durch die Interaktion des Spielers bestimmt.
- Moves: Die Möglichkeit, verschiedene Funktions-Sequenzen zu verschachteln und dabei eine Konsistenz der Geschichte aufrecht zu erhalten.

Die polymorphe Funktion entspricht im Wesentlichen einer Übertragung des Konzept der polymorphic Beats²⁵ auf die höhere Abstraktionsstufe der Funktion. Um den Zusammenhang zwischen Varianten und polymorphen Funktionen zu verdeutlichen wird dieser an einem Beispiel erklärt:

Der Anschlag eines Feindes gelingt in der klassischen Funktion nach Propp (Funktion **A**, *Villainy*) immer. In einer interaktiven Geschichte kann der Benutzer die Möglichkeit haben, den Anschlag des Feindes zu vereiteln. Nun kann die Funktion **A** in eine polymorphe Funktion A_{in} und in variante Funktionen A_{out} zerlegt werden. A_{in} - die Vorbereitung des Anschlages - gibt als Funktionswert das Gelingen oder Misslingen des Anschlages zurück. A_{out} besitzt Varianten für einen gelungenen oder mißglückten Anschlag des Feindes.

Tatsächlich kann Polymorphie und Varianten in einer Funktion vereinigt werden - so kann z.B. für die Propp-Funktion E, welche die Reaktion des Protagonisten auf einen Test des Wohltäters darstellt, direkt als Resultat angegeben werden, ob der Test bestanden ist oder nicht.

²⁴ Der Begriff der Funktion ist in diesem Zusammenhang ähnlich dem Klassenbegriff zu werten. Jede Klasse wird irgendwann instanziiert, die Instanziierung der Funktion erfolgt durch die Szene.

²⁵ Vgl. 3.2.3.

Aus dramaturgischen Gründen kann es notwendig sein, dass bestimmte Story-Elemente tatsächlich geschehen. So ist es erforderlich, einen Grund für Suche des Helden zu generieren oder einen für den weiteren Verlauf der Geschichte notwendigen magischen Helfer zu übergeben. Dies wird in der Story Engine durch eine Schleife geregelt, die entsprechende Funktion bis zur Erfüllung ausführt oder nach einer gewissen Zeit eine sogenannte spontane Variante einer Funktion ausführt. Diese Variante erzwingt quasi das für die Geschichte notwendige Ereignis.

Der Move ist ein Konzept, um verschachtelte, interaktive Storys zu ermöglichen. So ist es z.B. in Propps Ansatz eigentlich nicht möglich, sowohl einen Kampf zu führen als auch ein Rätsel zu lösen. Mittels des Move-Konzeptes werden bestimmte zusammenhängende Sequenzen so abgearbeitet, dass der Zusammenhang²⁶ tatsächlich auch erhalten bleibt und die einzelnen Funktionen sich nicht gegenseitig stören. Folgende Kontextinformationen bestimmen den Zusammenhang:

on_stage(Figur1) Nur Figuren, die ungebunden Handeln können. Alle Figuren die in einer Szene vorkommen müssen theoretisch derart verkettet werden. Ausnahmen sind Statisten die in einer Szene auf und abtreten.

helfer(helfer1) helfer1 befindet sich im Besitz des Helden

zeichen(zeichen1) zeichen1 ist ein Erkennungszeichen des Helden.

risiko(ereigniss1) niedere Verkettung von Szenen. Dies muss bei Übergängen für einwandfreien Kontext sorgen.

unglück(ereigniss1) Gibt das geschehene Unglück an, das den Move dominiert, nur solange das Unglück noch besteht.

hintergrund(ereigniss1) Ein Abgelaufenes Unglück wechselt in den Hintergrund, was es als geschehen markiert.

Weitere Parameter, welche die Kontextinformationen ergänzen, sind absolut beliebig und werden vom Autor definiert.

5.3.2 Szenen

Eine Szene stellt eine Auswahl von geschichts-relevanten Ereignissen dar, die durch den Autor der Szene (und damit i.A. dem Autoren der Geschichte) getroffen wird. Zur Szenenwahl

²⁶ Moves dienen zur Verifikation des Kontextes einer Szene. Als Kontext wird dabei alles angesehen, was in der Szene an relevanten Objekten vorhanden sein muss beziehungsweise nicht vorhanden sein darf - zum Beispiel darf ein Darsteller, der in der Szene handelt, nicht im Kontext einer vorherigen Szene entfernt worden sein. Ist die Engine bezüglich einer Szene nicht in der Lage, diese Szene aufzulösen, daher bis zur End-Szene zu gelangen, so darf diese Szene nicht durch die Story Engine ausgewählt werden. Dieses Konzept ist insbesondere bei Video-Clips sehr wichtig, da entsprechende Darsteller in verschiedenen Clips notwendigerweise übereinstimmen müssen oder die Darstellung durch den synthetischen Actor übernommen werden muss.

benötigt die Story Engine neben der Angabe, ob die Szene schon abgespielt wurde, weitere Kriterien, um eine Funktion auf eine Szene abbilden zu können bzw. um Sequenzen von Funktionen berechnen zu können:

- Funktionelle Einschränkungen: Eine Szene kann i.A. nur für bestimmte Funktionen als Instanziierung herangezogen werden. Eine Sequenz von Funktionen ist somit nur dann gültig, wenn entsprechende Instanziierungen der Funktionen möglich sind.
- Zeitliche Limitation: Um einen eventuellen Zeitrahmen einzuhalten sind die Szenen mit einer minimalen und einer maximalen Spieldauer annotiert. Je nach Spielzeit einer Szene (welche eventuell direkt durch den Benutzer beeinflusst ist) verbleibt eine restliche Dauer der Geschichte. Eine Sequenz von Funktionen ist nur dann gültig, wenn zumindest eine mögliche Instanziierung der Funktionen existiert, welche die Gesamtdauer der Geschichte nicht überschreitet.
- Kontext der Geschichte: Der Kontext einer Szene besteht z.B. aus Charakteren, welche zur Durchführung der Szene benötigt werden oder Aktionen (z.B. Anschläge, Unglücke), die Voraussetzung für eine Szene sind. Der Kontext der Szene muss zum Szenenbeginn mit dem Gesamtkontext der Geschichte konform sein. Natürlich kann innerhalb einer Szene der Gesamtkontext verändert werden - so können z.B. weitere Charaktere eingeführt werden.
- Benutzermodell: Das Benutzermodell der Szene muss konform zum Benutzermodell der Gesamtgeschichte sein. So wird z.B. verhindert, dass eine bestimmte Altersgruppenangabe (z.B. *unter 16 Jahre*) nicht durch die Szene verletzt wird (z.B. durch eine Altersfreigabe *ab 18 Jahre*).
- Dramaturgische Einschränkungen: Eine Szene kann als besonders schwierig oder als besonders spannend gekennzeichnet werden, um entsprechende Anpassungen bezüglich der Belastung des Benutzers vornehmen zu können.

Die funktionale Interpretation von Szenen als morphologischer Anteil einer Geschichte für die Zusammenstellung von nichtlinearen Geschichten auf Basis von Video-Clips entspricht der selektiven Komponente des Konzeptes. Die generative Komponente wird im folgenden Kapitel diskutiert.

5.3.3 Narrativer Einsatz von Konversation

Der Konversationsaspekt des Systems wird in direktem Zusammenhang mit den generischen Komponenten des Systems gesehen. Erweitert ein System den narrativen Ansatz um eine generische Komponente wie einen menschähnlichen Avatar bzw. synthetischen Actor, so kann diese Komponente in verschiedenen, in Tabelle 5.2 beschriebenen, narrativen Weisen eingesetzt werden, vgl. Braun [Bra01a]. Die drei möglichen Einsatzformen des Avatars werden alle durch die vorgegebenen Kontexte - die Inhalte der Geschichte - gesteuert. Die Nutzung ist somit zentriert auf das darzustellende Medium. Insbesondere wird durch den

Tab. 5.2: Narrative Einsatzweise eines konversationalen Avatars

Einsatz	Beschreibung
Teil der Geschichte	Der Avatar wird als narrativer Teil der Erzählung genutzt. Die Restriktionen, die sich aus der Nutzung von Video-Clips als selektive Grundlage der Geschichte ergeben (wie zum Beispiel feste Charaktere, Aktionen, Spannung) werden beim Avatar umgangen durch das Einsetzen eines generativen Verhaltens. Als Resultat dieser erweiterten Erzählstruktur wird der Avatar durch den Benutzer als (mimetischer) Teil der Geschichte wahrgenommen - als <i>innerhalb</i> der Storywelt.
Konferencier	Der Avatar wird als Showmaster beziehungsweise als Diskussionsleiter der Geschichte genutzt. Im Unterschied zum vorherigen Punkt agiert der Avatar außerhalb der Storywelt. Dadurch ist er für den Benutzer als separates Individuum außerhalb der Geschichte zu erkennen - jedoch mit einem z.B. navigatorischen (diegetischen) Einfluss auf die Geschichte.
Publikum (Audience)	Der Avatar wird als Zuschauer genutzt, dies bedeutet er zeigt Verhalten und Sprache als ob er ein Mit-Zuschauer ist. Im Gegensatz zu den vorherigen Punkten hat der Avatar für den Zuschauer keinen merkbaren Einfluss auf die Narration der Geschichte. Der Avatar wird natürlich trotzdem für die Geschichte genutzt - er forciert Emotionen und Reaktionen des Publikums, insofern ist er aus Sicht eines Geschichts-Autoren sehr wohl als Teil der Geschichte anzusehen.

narrativen Einsatz des Avatars die konversationale Einflussnahme des Publikums auf die Geschichte personifiziert und damit in einen sozialen, menschähnlichen Rahmen gesetzt.

Der narrative Effekt der konversationalen Komponente und des Avatars als generative Komponente an sich wird auch durch die Kameraeinstellung des Avatars beeinflusst. Eine Automatisierung der, aus dem Bereich Film bekannten, Kameraführung für nichtlineares Storytelling ist in der Literatur als Camera Planning bekannt. Die Kameraführung wird dabei meist als Benutzeraktion angeboten - und dient in diesen Fällen nicht zur Einschränkung von Interaktionswünschen des Benutzers.

Bates, vgl. [BS89], fordert eine emotional-interaktive Komponente und beschreibt, wie die grundlegende Filmtechnik darum erweitert werden kann. Das Ausnutzen der in den grundlegenden Filmtechniken erkannten emotionalen Trigger des Benutzers soll ihn in einer Situation für spezielle Interaktionen empfänglich machen; für alle anderen in der Situation grundsätzlich denkbaren Interaktionen jedoch unempfänglich. Dieser Ansatz ist prinzipiell konform mit dem Ansatz von Spierling und Gaida, vgl. [SG00], der die Ausnutzung von Kameraperspektiven beschreibt (welche im Kapitel 3.1.2 erläutert sind), dies zur vereinfachten Interaktion mit einem natursprachlichen, virtuellen Repräsentanten in Form eines Avatars, siehe Abbildung 5.5. Der Avatar besitzt im vorgestellten Konzept dieser Arbeit folgende narrative Ausdrucksweisen:

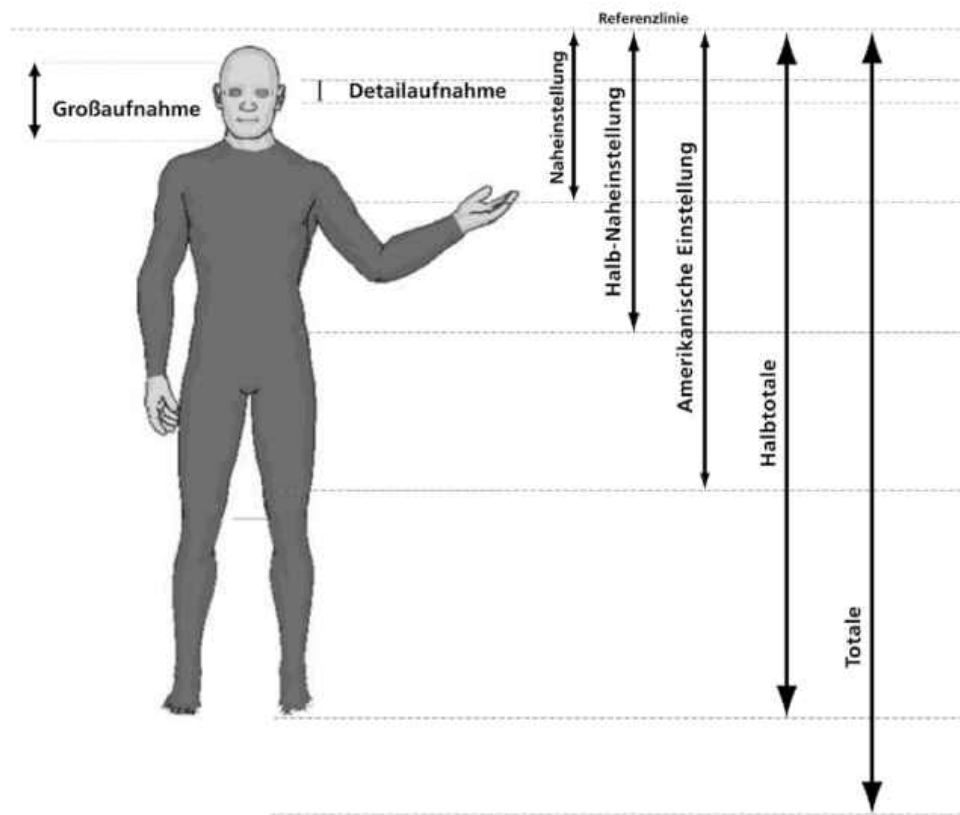


Fig. 5.5: Kameraeinstellungen

- Sichtweise des Avatars:

1. Der Avatar kann begleitend zur Information neben derselben dargestellt werden und ist somit gestalterisch und emotional ein separates kommunikatives Element der Präsentation, siehe Abbildung 5.6.



Fig. 5.6: Avatardarstellung separat zur Informationsdarstellung

2. Der Avatar kann als über die Videoinformation gelegter Teil bzw. in das Video integrierter, separat zu animierender Teil der Präsentation gesehen werden, siehe Abbildung 5.7..

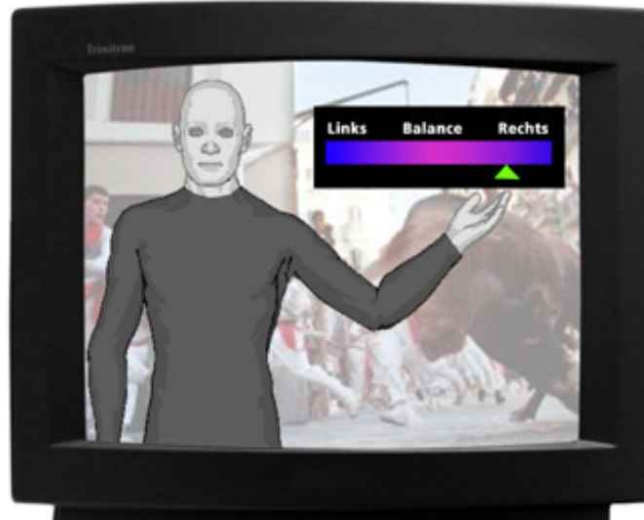


Fig. 5.7: Avatardarstellung integriert in Informationsdarstellung

- Der Avatar kann zur konversationalen Interaktionsführung und Präsentation von Video- und Textmaterial genutzt werden.
- Eine emotionale Ausdrucksfähigkeit ist dem Avatars möglich.

Als geeignete Kameraeinstellungen auf den Avatar werden, unabhängig von der Fallunterscheidung der Sichtweise, die Großaufnahme, die Naheinstellung, die Halbnaheinstellung, die 'Amerikanische' und die Halbtotale identifiziert. Mit diesen Einstellungen kann der Avatar und somit die konversationale Komponente des Konzeptes als narrativer Bestandteil des Konzeptes genutzt werden.

5.4 Zusammenfassung

Dieses Kapitel stellt ein Konzept zur interaktiven Narration von kontinuierlichen Medien am Beispiel von Video vor. Das Konzept bietet eine Trennung von Erzähler (Narration), eigener Handlung des Publikums (direkt manipulative Interaktion) und konversationaler Beeinflussung des Erzählers (Konversationsmodellierung). Die allgemein zu modellierenden Anteile des Konzeptes sind amodal formuliert, d.h. ohne einen direkten Bezug zur medialen Ausprägung der Geschichten, was eine hohe Wiederverwendbarkeit der Komponenten impliziert. Die direkt manipulativen Anteile des Konzeptes müssen eine mediale Ausprägung besitzen, sie sind für das Medium Video definiert. Die einzelnen Eigenschaften sind wie folgt notiert:

- **Nichtlineare Narration:** Das Geschichtenmodell ist unabhängig von dem Geschichtenerzähler (Story Engine). Der Geschichtenerzähler erzählt die Geschichte aufgrund einer morphologischen Zuordnung von einzelnen Geschichtsfragmenten, welche in Abhängigkeit von der morphologisch ausgewerteten Interaktion des Benutzers (geschichtsspezifische Storyacts) zu einer Geschichte zusammengestellt werden. Diese Methode bietet verschiedene Freiheitsgrade der Interaktion durch den Benutzer, einen kontrollierten Spannungsaufbau nach einem Geschichtenmodell, eine Abgeschlossenheit der Geschichte auf Zeit- und Handlungsebene, eine Adaptierbarkeit auf den Benutzer sowie eine explizite Zugänglichkeit durch Autoren. Da das morphologische Storymodell im Abstraktionsgrad über der szenischen Beschreibung angeordnet ist, kann das Storymodell direkt auf die Narration mit szenisch nicht zu beeinflussendem, selektivem Material wie Video-Clips angewendet werden.
- **Konversationale Interaktion:** Die Modellierung der Konversation geschieht explizit und unabhängig von der verwendeten Modalität der Inhaltsdarbietung durch die Geschichte. Die verschiedenen Aspekte von Konversation - sowohl propositionale als auch diskursbezogene Komponenten - werden zu abstrakten Szenarien notiert und mittels eines Regelsystems, erstellt durch menschliche Experten, manipuliert.
- **Medienspezifische Contentprovider** bieten einen - im vorliegenden Falle auf Video - optimierten, direkt manipulativen Zugriff auf die Geschichte. Für Video wird eine explizite, temporale (explizite Darstellung der zeitlichen Entwicklung der Manipulationsmöglichkeit), und intramediale (Manipulation im gleichen Medienkanal) Annotation für Video konzipiert.

Das Konzept impliziert folgende Vorteile:

- Die explizite konversationale Modellierung schafft einen einfach zu modifizierenden, designspezifischen Ansatz, der expertenorientiert arbeitet. Durch die nicht medienspezifische Notation ist der Ansatz generisch und auf verschiedenste Applikationen übertragbar.
- Nichtlineare Narration erlaubt eine explizite Kontrolle der Geschichtsstruktur und die Vorgabe eines Geschichtsziels (explizit erreichbares Ende der Geschichte), ohne das der Benutzer tatsächlich in seiner Handlungsfreiheit störend eingeschränkt wird.
- Der applikationsspezifische Einsatz von direkt manipulativen Elementen erlaubt eine Optimierung derselben auf das verwendete Medium, was eine einfache und intuitive Bedienung impliziert. Die eigene Handlung des Publikums in Bezug zur Story ist medienspezifisch, da sehr große Unterschiede zwischen z.B. Virtual-Reality-Szenen und interaktiven Video-Clips.

Das vorgestellte Konzept dient als Basis zur Prüfung der in Kapitel 1.1 aufgestellten Arbeitshypothesen. In Kapitel 6 wird die Umsetzung des Konzeptes in drei Applikationen validiert, in Kapitel 7 werden die Resultate der Umsetzung in Bezug auf die durch das Konzept implizierten Erwartungen und damit die Arbeitshypothesen diskutiert.

6. SYSTEMVALIDIERUNG

Dieses Kapitel beschreibt Applikationen, die das spezifizierte Systemkonzept zum Teil oder gänzlich implementieren. Anhand der verschiedenen Applikationen wird die Effizienz und Güte des Ansatzes validiert. So fließt das Konzept dieser Arbeit wie im folgenden skizziert in die unten genannten Projekte und Studien ein:

EMBASSI Das BMBF Projekt EMBASSI (**E**lektronische **M**ultimediale **B**edien- und **S**ervice **A**ssistenz) implementiert die Charakter bezogenen Anteile des Konzeptes. Weiterhin implementiert EMBASSI die konversationalen Anteile des Konzeptes, allerdings nicht in einer expliziten Notation des Konversationsmodells, wie sie im Konzept gefordert ist. Die kontinuierliche Medienpräsentation in Kombination mit direkt manipulativer Interaktion wird im EMBASSI Projekt ebenfalls implementiert, allerdings fließen die geforderten Konzepte zur temporalen, expliziten, intramedialen Interaktion auf den Medien sowie einer nichtlinearen Story Modellierung nicht im EMBASSI Projekt ein.

MAP Das BMBF Projekt MAP (**M**ultimedia **A**rbeitsplatz der Zukunft) implementiert eine explizit modellierte konversationale Interaktion, angetrieben durch Kontext-vorgebende Applikationen in Form von Agenten. Im MAP Projekt sind, wie im Konzept dieser Arbeit gefordert, die modalen Teile der Konversation von der amodalen Diskursmodellierung getrennt, für beide Teile wird jeweils eine eigene Modellierung angegeben. Direkt manipulative Interaktionsmöglichkeiten werden durch die Agenten selbst realisiert, so wird u.a. ein Agent zur interaktiven Darbietung von Videokonferenzen angeboten - dieses allerdings ohne den Einsatz einer temporalen, expliziten, intramedialen Interaktion oder eines nichtlinearen Story Modells.

DIVA Die Studie DIVA (**D**igital **I**nteractive **V**ideo and **A**udio) implementiert den im Konzept für interaktives Video geforderten kombiniert generativ/selektiven Ansatz der Narration sowie eine kombinierte konversationale und direkt manipulative Interaktionsmöglichkeit auf dem Medium. Die konversationale Interaktionsmöglichkeit ist mit einem anthropomorphen Ansprechpartner modelliert. Dieser kann separat durch die kontext-gebende Videoapplikation angesprochen werden. Die kontinuierliche Medienpräsentation geschieht mit einer temporalen, expliziten und intramedialen Interaktionsmöglichkeit. Die Videopräsentation selbst ist nichtlinear und wird über ein Story Modell kombiniert selektiv (mittels Video-Clips) und generativ (mittels eines animierten Avatars) aufgebaut.

Mit diesen Ansätzen wird nicht zwingend die Implementierung des Gesamtkonzeptes angestrebt - eine solche ist zweifelsfrei machbar, jedoch an dieser Stelle nicht sinnvoll, da sie nicht zeigen würde, wie Teile oder das gesamte Konzept zur Lösung von konkreten Anwendungsanforderungen und Anwendungsproblemstellungen in verschiedenen Szenarien genutzt werden. Die Szenarien der drei Anwendungen unterscheiden sich dabei vor allem in der Anwendungsbreite der Gebiete - so sind EMBASSI und MAP sehr allgemein gehalten, generische Ansätze, während das DIVA System eine spezialisierte, jedoch ebenfalls generische Anwendung ist.

Im Folgenden werden die Projekte EMBASSI, MAP und die Studie DIVA detailliert in Bezug auf das in Kapitel 5 vorgestellte Konzept diskutiert. Im Anschluss wird zusammenfassend auf die Konzeptumsetzung eingegangen.

6.1 EMBASSI

Das EMBASSI-Projekt - Elektronische Multimediale Bedien- und Service Assistenz - startet als ein auf 4 Jahre ausgelegtes Leitprojekt vom Bundesministerium für Bildung und Forschung (BMBF) innerhalb des strategischen Forschungsbereichs Mensch-Technik-Interaktion. 19 Partner aus Industrie, außeruniversitärer Forschung und Hochschulen erarbeiten einen integrierten Ansatz für die Entwicklung von Assistenten für Alltagstechnologien.

Die Grundlegende Ausrichtung des EMBASSI Projekts orientiert sich auf den Übergang von einer funktionen-orientierten auf die zielorientierte Interaktion mit Systemen - daher das Ersetzen der Werkzeug-Methapher durch eine Assistenzmetapher - und den Übergang von unimodalen, menübasierten Dialogstrukturen (WIMP-Interfaces) zu multimodalen Konversations- und Dialogstrukturen. Die folgenden Aspekte sind dabei berücksichtigt, siehe [HK01]:

- Die Definition eines einheitlichen, psychologisch fundierten Vorgehensmodells zur systematischen Entwicklung von Assistenzsystemen mit multimodalen Nutzer-Schnittstellen.
- Die Entwicklung einer Design-Umgebung, die auf dem oben genannten Vorgehensmodell aufbaut und Hintergrundwissen sowie die Unterstützung von Expertensystemen für Systemdesigner zur Verfügung stellt.
- Die Entwicklung eines generischen Architekturrahmens mit einem angemessenen Satz von Protokollen für die Verwirklichung von Assistenzsystemen, die auf interoperablen Komponenten basieren.
- Die Entwicklung eines Technologiebaukastens, der aus modularen Bausteinen für multimodale Interaktions- und Assistenzfunktionalitäten besteht.

EMBASSI generiert somit in erster Linie eine Infrastruktur für nicht-berufliche Anwendungsszenarien aus den Bereichen Automobil, öffentliche Terminalsysteme und Haushalt.

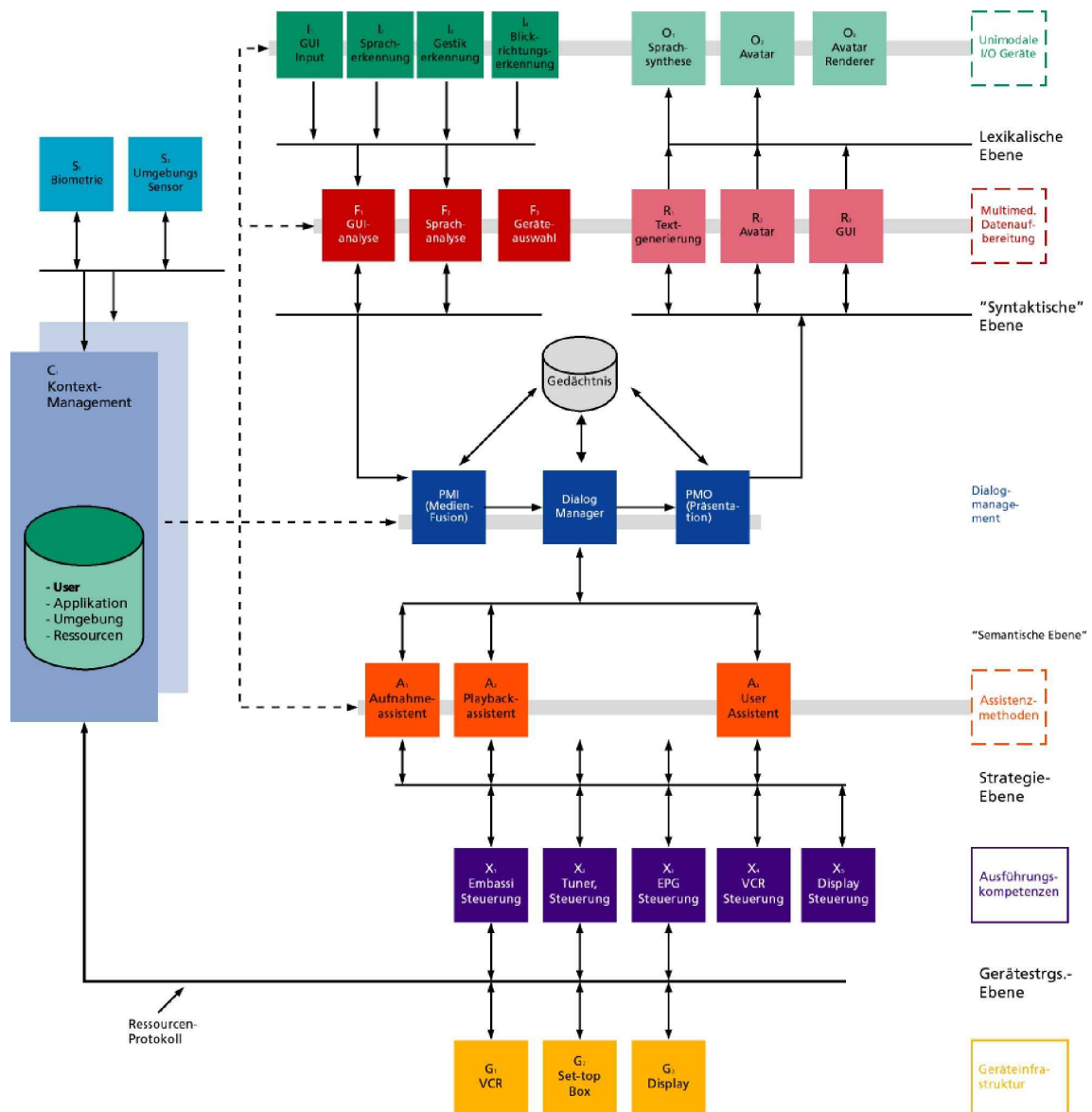


Fig. 6.1: EMBASSI Systemarchitektur, [HK01]

Der Begriff des Agenten entspricht im EMBASSI Projekt dem Assistenz-Begriff, wie er in Kapitel 2.2 definiert ist.

Die Phase I des zwei-phasigen Projektes ist inzwischen abgeschlossen. In diesem Zeitraum wurden individuelle, prototypische Systeme für die oben genannten Anwendungsszenarien entwickelt. Die EMBASSI Systemarchitektur ist in Abbildung 6.1 dargestellt. Im EMBASSI Projekt ist eine zentrale Komponente zur Steuerung der Dialoge - ob direkt manipulativ oder konversational - realisiert. Diese Komponente wird als *Dialogmanagement* (im engen

Zusammenspiel mit dem *Kontext-Management*) bezeichnet. Der Dialogmanager realisiert auf semantischer Ebene des EMBASSI Systems die Kommunikation zwischen Geräten sowie zwischen Geräte-spezifischen Agenten auf strategischer und Geräte-spezifischer Ebene des EMBASSI Systems, sowie den Eingabe- und Ausgabekomponenten (auf der syntaktischen und lexikalischen Ebene des EMBASSI Systems). Die Ein- und Ausgabe von semantischer Information, d.h. die Wandlung von modaler Information der Eingabegeräte in abstrakt-symbolische Information, interpretierbar durch den Dialogmanager, ist durch einen sogenannten *Polymodalen*¹ *Inputmanager* realisiert. Die Transformation von abstrakt-symbolischer Ausgabe-Information in modale Ausgaben wird durch einen sogenannten *Polymodalen Outputmanager* realisiert.

6.1.1 Dialogmanagement

Das Anwendungsszenario Privathaushalt A-PHH ist mit einer Vielfalt von beteiligten Komponenten bestückt, welche sich durch eine ausgeprägte Heterogenität auszeichnen. Um dem vielfältigen Funktionsumfang Herr zu werden wird in EMBASSI ein offenes System der Bedienassistent mit einer homogenen Bedienphilosophie realisiert, die unter anderem eine Nutzung der Verflechtung und Dynamik zwischen den Geräten ermöglicht.

Die Bedienassistent in A-PHH soll dabei auf einem Charakter basieren, jedoch verschiedenste Assistentenfunktionen, vgl. [BEG⁺00], ermöglichen, wie zum Beispiel:

- TV-Programmagent
- Aufnahme/Wiedergabe Programmier Assistent
- Bedienassistent im graphischen User Interface
- Shoppingassistent
- Hilfeassistent

Die Assistentenfunktionen werden - abhängig von der Präferenz des Benutzers - im EMBASSI System über verschiedene Modalitäten angesprochen. So kann sowohl eine Graphische Benutzeroberfläche als auch ein Avatarsystem zum Einsatz kommen.

Um eine konversationale Interaktion zwischen System und Benutzer zu ermöglichen ist eine auf Performatives [LF94] aufbauende Kommunikation zwischen Assistenten und Ein/Ausgabe-Komponenten, organisiert durch einen Dialogmanager, vgl. [AES⁺00], realisiert.

Performatives beschreiben die zwischen Agenten zulässigen Operationen, um die jeweiligen Eigenschaften eines Agenten nutzen zu können. Dieser Ansatz erweitert die Performatives auf die logische Beschreibung der Assistentenfunktion und ihrer Umsetzung durch die verschiedenen möglichen User Interface Komponenten, vgl. [EM00].

Durch den Dialogmanager werden Turntaking Signale (ToSystem, ToUser) ebenso aufgenommen wie aus SpeechActs, vgl. [Kle00], abgeleitete konversationale Anweisungen. In einer

¹ Der Ausdruck *polymodal* entspricht dem Ausdruck *multimodal*.

an Description Logic angelehnten Syntax werden diese konversationalen Aspekte kombiniert mit den tatsächlichen Systemzielen (d.h. dem darzustellenden Kontext). Die Kommunikation mit den Ein-/Ausgabekomponenten erfolgt über einen Polymodalen Inputmanager (PMI) und einen Polymodalen Outputmanager (PMO).

Die Dialogsteuerung bzw. die Kommunikation im EMBASSI System wird von den drei Komponenten Dialogmanager, PMI und PMO kooperativ umgesetzt:

Eingabe Die sogenannte Polymodale Eingabekomponente (PMI), vgl. [SR00], übernimmt die Fusion der unimodalen Eingaben wie Blickrichtungserkennung, Gestikerkennung und Spracheingabe zu einer multimodalen Eingabe, die im Anschluss auf eine amodale Eingabebeschreibung des Benutzers abgebildet wird. Die Aufgaben der PMI bestehen hauptsächlich in der Auflösung von Mehrdeutigkeiten unimodaler Eingaben durch Ausnutzen von zeitlichen Abhängigkeiten und syntaktischen Kenntnissen aufgrund von dem System bekannten Kommunikations- und Interaktionsstrukturen sowie Interaktions- und Sprachmodellen. Der PMI transferiert diese Informationen an den Dialogmanager.

Dialogmanager Der Dialogmanager erhält vom PMI eine amodale Eingabebeschreibung. Diese Eingabebeschreibung wird auf eine semantische Beschreibung des Benutzerzieles abgebildet. Die Benutzerziele werden in Aufträge für die jeweiligen Assistenten der syntaktischen Ebene, siehe Abbildung 6.1, umgesetzt. Die erreichten Ziele werden in einer amodalen Ausgabebeschreibung abgebildet, die der Polymodalen Ausgabekomponente zur Verfügung gestellt wird.

Ausgabe Die Polymodale Ausgabekomponente (PMO), vgl. [EMR⁺00], entscheidet die Modalitäten, welche bezüglich des Dialoges angesprochen werden sollen, indem sie die amodale Ausgabebeschreibung in einer multimodalen Ausgabe abbildet. Die verschiedenen unimodalen Ausgabekanäle werden direkt von der PMO mit entsprechenden Kommandos angesprochen und synchronisiert.

Der Dialogmanager ist direkt aus einer Natursprache-verarbeitenden und generierenden² Applikation abgeleitet. Die benutzten SpeechActs geben eine propositionale Sichtweise auf die zu bewältigenden Aufgaben des EMBASSI Systems wieder.

Die Synchronisation zwischen den verschiedenen Ausgabekomponenten erfolgt wie im folgenden Beschrieben:

1. Der PMO leitet alle synchron wiederzugebenden Aufgaben an die verschiedenen Ausgabekomponenten.
2. Der PMO weist explizit auf das Ausführen der Aufgaben hin.
3. Der PMO erwartet die Erfolgsmeldungen der Komponenten. Danach fährt er mit Schritt 1 fort.

² Vgl. Kapitel 4.1.1.

Die gesamte abstrakte Konversationsmodellierung findet in der Pipeline Polymodaler Inputmanager → Dialogmanager → Polymodaler Outputmanager statt, siehe auch Abbildung 6.1. Die modale Umsetzung der Konversationsausgaben des Systems findet direkt in den Ausgabekomponenten statt.

6.1.2 Konversationaler Charakter

Um die im EMBASSI-Kontext geforderte visuelle Präsentation und Assistenz mittels anthropomorpher Avatare beziehungsweise synthetischer Aktoren zu realisieren ist eine Avatar Plattform, bestehend aus einem Avatar Controller und zugehörigem Avatar Renderer, realisiert. Die Avatare dienen dabei der Realisierung von Assistenzfunktionen unter Einbeziehung von natürlichen, bidirektionalen Kommunikationsformen wie der Verwendung von Sprache, Mimik und Gestik sowie Posen.

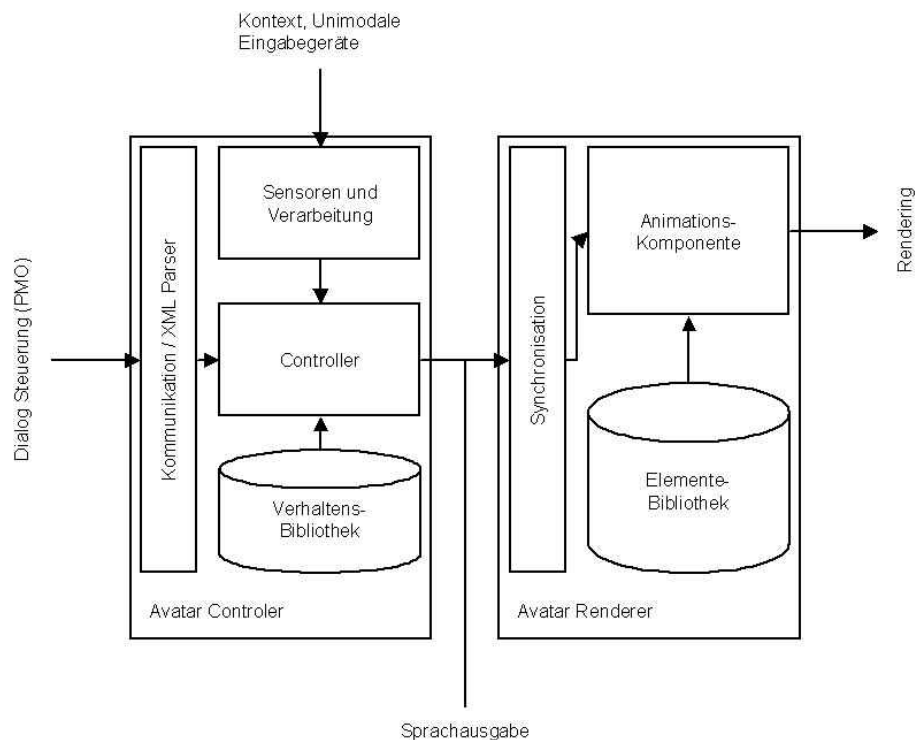


Fig. 6.2: EMBASSI Avatar Controller und Renderer

Die Avatar-Plattform ermöglicht hierbei dem EMBASSI Dialogmanager die Instanziierung und Ansteuerung eines Avatars: Wie im vorangegangenen Kapitel festgestellt wird die konversationale Modellierung des Dialoges zwischen EMBASSI System und Benutzer weitgehend durch den Dialogmanager und Polymodalen Outputmanager getätigt. Die entsprechende Umsetzung der abstrakten, konversationalen Anweisungen sowie eine Generierung von (zur Sprachausgabe) synchronen Lippenbewegungen (sogenannten Visemen, vgl.

[MAR⁺00b)) ist die Aufgabe der Avatar-Plattform, siehe Abbildung 6.2.

Tab. 6.1: Abstraktionsstufen der Avatar Steuerung

Ebene	Bedeutung
Geometrie	Diese Ebene bezeichnet die konkrete Manipulation von geometrischen Ausprägungen (Polygonen) des Avatar Modells.
Feature	Die Ebene der Eigenschaften ermöglicht die Ausgabe predefinierter Mimiken, Gestiken und Posen. Die Synchronisation mit Sprachausgabe wird dabei ebenfalls als Eigenschaft aufgefasst, d.h. entsprechende Ansteuerung von Visemen erfolgt auf dieser Ebene.
Task	Die Ebene der Aufgaben ermöglicht eine Ansteuerung, welche schon über eine konkrete zeitliche Ausdehnung - unabhängig von der zu synchronisierenden Sprache - verfügt. Aufgaben sind dabei definiert als vollständige Bewegungsabläufe und Verhaltensmuster, die für den Benutzer als in sich abgeschlossene Verhaltenseinheit zu erkennen sind.
Motivation	Die Ebene der Motivationen stellt ein - durch eine eigenständige Psyche und Emotionalität des Avatarmodells geprägte - Methode der Ansteuerung von Stimmungen und Einstellungen des Avatars wieder. Tatsächlich fungiert sie nicht als Ebene der expliziten Darstellung von konversationalem Verhalten. Die Motivations-Ebene ist unabhängig von der Task-Ebene zu sehen, jedoch wird bei der tatsächlichen Verhaltensbildung die Task Ebene von der Motivationsebene überlagert.

Mittels eines Avatar Controllers wird eine mehrstufig abstrakte Ansteuerung des Avatars, siehe Tabelle 6.1, zur Verfügung gestellt, die Avatarbefehle, siehe Tabelle 6.2, 6.3 und 6.4, in für den Avatar Renderer lesbare Feature-Sequenzen, vgl. [AM99], übersetzt.

Der Avatar Renderer, ein Java 3D basiertes Feature Morphing System, vgl. [AM99], synchronisiert eigens erzeugte Lippenbewegungen zur Sprachausgabe und rendert den Avatar auf entsprechendes Ausgabegerät wie zum Beispiel einen Fernseher.

Die Auswahl und Erzeugung eines Avatars geschieht auf Basis der integrierten Elemente Bibliothek, welche entsprechende Avatar Modelle gespeichert hat. Diese Avatare können auf Geometrischer- und auf Eigenschaften- (Feature-) Ebene angesteuert werden - dies impliziert zusätzlich die Synchronisation mit einer Sprachausgabe durch den Renderer, siehe Tabelle 6.1. Eine weitere zeitliche Kontrolle oder Synchronisation erfolgt jedoch nicht im Render.

Der Avatar Controller ermöglicht neben der Ansteuerung des Avatars auf Geometrie- und Eigenschaften-Ebene auch eine Ansteuerung auf abstrakten Ebenen wie der Aufgaben- (Task-) und Motivations-Ebene. Er verfügt über eine Verhaltensbibliothek, in welcher entsprechende Animationsstrukturen in Form von Befehlssequenzen abgelegt sind. Entspre-

Tab. 6.2: Avatar Befehle auf Motivation Ebene

Ausdruck	Befehl
SEI NORMAL	NORMAL
SEI FRÖHLICH	HAPPY
SEI AUFGEREGT	NERVOUS
SEI INTERESSIERT	INTERESTED
SEI DEMÜTIG	HUMBLE
SEI BESCHÄFTIGT	BUSSY
SEI SUCHEND	SEARCHING
SEI ÜBERFORDERT	SWAMPED
SEI UNTERFORDERT	DESWAMPED
SEI BELEHREND	INSTRUCTING
SEI ERKLÄREND	ILLUSTRATIVE
SEI ZUSTIMMEND	ASSENTED
SEI ABLEHNEND	DISSENTED
SEI NERVÖS	JITTERY
SEI ANBIETEND	OFFERING

chend den angeforderten Aufgaben können diese Strukturen aktiviert werden, wobei keine feste Reihenfolge vorgegeben wird und die Aufgaben durch den Avatar selbst auf Strukturen abgebildet werden - so kann aus verschiedenen möglichen Strukturen für Aufgaben frei ausgewählt werden, um eine gewisse Spontanität der Handlung zu gewährleisten.

Der Avatar-Controller empfängt seine auszuführenden Anweisungen auf der syntaktischen Ebene von der PMO:

- Dieser vermittelt dem Avatar-Controller mittels des Parameters *speechact* des Tags *DL* die Grundlage zur Verhaltensbestimmung. Vom AvatarController werden geeignete Verhaltensbausteine bereitgestellt, die die verschiedenen SpeechActs begleiten können.
- Um Gestik bzw. Blick/Kopfrichtung generieren zu können wird dem Avatar-Controller in der Komponente *Coordinates* die entsprechende Koordinate des Ziel-Objekts angezeigt. Die Turntaking-Informationen werden separat im Tag *Turntaking* mit dem Parameter *direction* an den Avatar-Controller übergeben.
- Alternativ kann die PMO auch aus einer vordefinierten Menge von Stimmungen (siehe Tabelle 6.2), Aktionen (siehe Tabelle 6.3, z.B. Nicken des Kopfes oder Heben der Augenbraue) oder gezielte Bewegungen (siehe Tabelle 6.4) auswählen. Jede Stimmung kann einer oder mehrerer momentan auszuführenden Aktionen überlagert werden, entsprechend überlagern Aktionen eine oder mehrere gezielte Bewegungen. Auf der Geometrieebene sind keine Modellierungsmöglichkeiten seitens des Dialog-Managers vorgesehen.

Tab. 6.3: Avatar Befehle auf Task Ebene

Ausdruck	Befehl
ZEIGEGESTE	SHOW
ÖFFNE	OPEN
SCHLIESSE	CLOSE
Turntaking: Gesprächsführung an Benutzer abgeben	TOUSER
Turntaking: System übernimmt Gesprächsführung	TOSYSTEM
Backchannel	Agreement
Backchannel	Confusion

Tab. 6.4: Avatar Befehle auf Feature Ebene

Ausdruck	Befehl
KOPF NICKEN	NOD_HEAD
KOPF SCHÜTTELN	SHAKE_HEAD
LINKE AUGENBRAUE NACH UNTEN	EYEBROW_LEFT_DOWN
LINKE AUGENBRAUE NACH OBEN	EYEBROW_LEFT_UP
RECHTE AUGENBRAUE NACH UNTEN	EYEBROW_RIGHT_DOWN
RECHTE AUGENBRAUE NACH OBEN	EYEBROW_RIGHT_UP

Die Befehle dienen dem Zweck, den allgemeinen temporal-spezifischen Zustand des Avatars in Bezug auf die Benutzerinteraktion zu spezifizieren. Gewisse Grundzustände wie *WÜTEND* sind im Bezug auf die Benutzerinteraktion wenig sinnvoll, da das System eine Assistenz verkörpert und keinen Gegner des Benutzers darstellt. Der Motivation-Level wird somit durch eine emotional-sachliche Grundstimmung geprägt.

Die entsprechenden Befehle werden in eine XML Struktur eingebettet, mit welcher dem Avatar Controller die Dialog-Anforderungen an den Avatar übermittelt werden können.

6.1.3 Konzeptumsetzung

Das EMBASSI System setzt das in Kapitel 5 geforderte Konzept in Teilbereichen um, da die applikationsspezifischen Anforderungen im EMBASSI Projekt vor allem die Entwicklung einer plattformübergreifenden, einheitlichen Assistenzmöglichkeit vorsehen und die nichtlineare Narration kontinuierlicher Medien nur eine Randbedingung der verschiedenen Partner ist. So sind die drei Bereiche Narration, Direkt Manipulative Interaktion und Konversationale Interaktion nicht getrennt - wie Abbildung 6.1 zeigt wird die gesamte

Kommunikation zwischen Benutzer und Geräten von einer zentralen Komponente übernommen - dem Dialogmanager. Eine Trennung von konversationalem und direkt manipulativem Konzept erfolgt nicht - der Dialogmanager spricht über den Polymodalen Outputmanager sowohl direkt manipulative als auch konversationale Komponenten des Benutzerinterfaces an.

Im Einzelnen können die folgenden Punkte an der Umsetzung des konversationalen Konzepts kritisiert werden:

- Der Dialogmanager leitet die konversationale Interaktionsmöglichkeit des Benutzers direkt von Methodiken der Sprachanalyse und Sprachausgabe ab - er ist somit optimiert auf natursprachlich basierte Kommunikation, eine Berücksichtigung von visuellem Kommunikationsverhalten wie Mimiken, Gestiken und Posen findet in der konversationalen Modellierung des Dialogmanagers keinen Raum - eine entsprechende Annotation der konversationalen Inhalte findet erst in der Ausgabekomponente Avatar, der Avatarplattform, statt.
- Das Dialogkonzept des Dialogmanagers verwendet keine explizite Notation der konversationalen Aspekte einer Konversation - diese Aspekte sind implizit in der aussagenlogischen Formulierung der Systemziele und der Annotation dieser Systemziele mittels Speech Acts inbegriffen. Das macht eine Änderung der konversationalen Modellierung des Systems zu einer verteilten Aufgabe: Diverse Komponenten des Systemes (Dialogmanager, Sprachausgabe, Polymodaler Input/Outputmanager, Avatarsystem, GUI-System), die keine expliziten Angaben bezüglich der konversationalen Aspekte der Interaktion mit dem Benutzer verstehen, werden in kleinen Teilen modifiziert, um eine Änderung der impliziten konversationalen Modellierung zu erreichen.
- Der Dialogmanager - als zentrale Komponente des Systems - ist daraufhin optimiert, sowohl direkte Manipulation als auch Konversation mit einem Satz von Befehlen - Speech Acts - zu kommunizieren. Dies muss zwingend dazu führen, dass die speziellen Eigenschaften der beiden verschiedenen Konzepte nicht voneinander getrennt bearbeitet werden - eine Vermischung des Konzepte wird die Benutzung der jeweiligen Konzepte durch den Anwendungsentwickler erschweren.

Das Konzept bezüglich der nichtlinearen Narration von kontinuierlichen Medien, realisiert durch Video, ist im EMBASSI Projekt nur rudimentär umgesetzt; Jedem Inhaltsproduzierenden Partner obliegt es selbst, ein entsprechendes Konzept einzusetzen. Im Wesentlichen ist im Bereich A-PHH des EMBASSI Projekts eine Unterstützung der linearen Narration von Video gegeben.

Die zentralisierte Informationsverarbeitung des EMBASSI Systems führt zu einer besonders schwierigen Situation für den Anwendungsentwickler - spezifische, direkt manipulative Zugriffe auf die angeschlossenen Geräte gestalten sich für ihn als unmöglich, da ein solcher Zugriff durch das System nicht vorgesehen ist. Der Entwickler muss jedwede Kommunikation dem Dialogmanager überlassen. Diese Tatsache wird jedoch durch den Anwendungsbereich, speziell im privaten Haushalt und generell im privaten Alltagsbereich entschärft, welcher

neben der Ein/Ausgabe mittels GUI Elementen nur wenige direkt manipulative Eingriffe des Benutzers erfordert.

Trotz der nur teilweisen Umsetzung des Konzepts für das EMBASSI System kann in Kapitel 7 gezeigt werden, dass der Anspruch einer konversationalen Benutzerschnittstelle zur Verbesserung der Assistenz durch das System schon mit dem EMBASSI Ansatz nachgewiesen werden kann.

6.2 MAP

Das MAP Projekt - Multimedia Arbeitsplatz der Zukunft - startete als Leitprojekt des Bundesministeriums für Wirtschaft und Technologie im Themenfeld Mensch-Technik-Interaktion in der Wissensgesellschaft, als Konsortium besteht aus 15 Partnern, im Sommer 2000 ein drei Jahre andauerndes Forschungsvorhaben, welches eine Plattform zur Kommunikation und Zusammenarbeit verschiedenster Agenten-Applikationen zum Ziel hat. Folgende Faktoren sind dabei die Forschungsschwerpunkte:

- Die Einbindung von mobilen Geräten mit intuitiv-multimodaler, interaktiver Bedienung von multimedialen Applikationen.
- Agentenbasierte Applikationsentwicklung mit sicherer Datenhaltung und Verarbeitung in verteilten Netzwerken.
- Nachweis der Benutzbarkeit durch die Implementierung einer entsprechenden Basisapplikation und dem Test der Applikation in einem (Bau)-spezifischen Szenario.

Der grundlegende Anspruch des MAP Systems lautet dabei: personenbezogene, menschenorientierte Assistenz und Delegation wo und wann immer der Benutzer es möchte. Dieser Anspruch folgt direkt aus dem beobachteten, nun anstehenden Paradigmenwechsel bei Arbeitsplatzsystemen, vgl. [WKL01], weg von den netzwerkbezogenen Einzelanwendungen hin zu Assistenz- und Delegations-Gruppenanwendungen in integrierten Netzwerken, siehe Bild 6.3.

Der Fokus des MAP Systems, siehe [WKL01], ist damit die multimediale Interaktion zur Delegation und Assistenz von Aufgaben durch den Benutzer an intelligente Systeme, dies durch das Bereitstellen von entsprechenden Interaktions- Technologien, Komponenten und Methoden. In einer Pilotimplementierung verrichten Software Agenten routinemäßige Aufgaben des Benutzers um diesen zu entlasten. Die Delegation sowie die Ergebnisdarstellung geschieht dabei in adaptierter, multimedialer Weise, dies wird in Abbildung 6.4 skizziert. Das System wird psychologisch und soziologisch getestet, um eine entsprechende Nutzbarkeit tatsächlich zu überprüfen und Richtwerte für Effektivität und Nutzerfreundlichkeit des Systems zu erhalten.



Fig. 6.3: Zeitliche Entwicklung der Arbeits- und Nutzungsformen an Rechnerarbeitsplätzen, [KP01]



Fig. 6.4: MAP User Interface Agent - Kommunikation zwischen Benutzer und Basisagenten

Das MAP System implementiert das Konzept³ für den Bereich konversationale Benutzerschnittstellen. Durch die MAP gegebene Struktur der parallelen *Basic Services*⁴ wird der Ansatz der applikationsspezifischen Konversations-Inhalte in Kombination mit applikationsspezifischer, direkt manipulativer kontinuierlicher Inhaltspräsentation verwirklicht⁵. Das MAP System wird im weiteren detailliert vorgestellt, anschließend auf die Konzept-realisierende Teilkomponente User Interface Agent (UIA-Agent) des MAP Systems eingegangen.

6.2.1 Das MAP System

Das MAP System versucht, den Anspruch der allgegenwärtigen Assistenz und Delegation mittels menschähnlichen Interaktionsmetaphern durch eine entsprechende Infrastruktur und eine generische, prototypische Implementierung zu unterstützen. Um diese Herausfor-

³ Vgl. Kapitel 5.2.2.

⁴ Vgl. Kapitel 6.2.3.

⁵ Vgl. Kapitel 5.1 und Kapitel 5.2.1.

derung tatsächlich zu realisieren sind verschiedene, wie im folgenden beschriebene, Schritte notwendig:

- Die Entwicklung von entsprechenden Sicherheitsmechanismen, die eine vertrauenswürdige Assistenz und Delegation in integrierten Netzwerken erst erlauben.
- Die Integration von virtueller Realität und allgegenwärtigem Computing in einem hybriden System.
- Die Entwicklung von Delegationsmechanismen unter Nutzung eines Agentensystems.
- Neue Interfaces, die verschiedene Modalitäten auf Ein- und Ausgabeseite nutzen, wie zum Beispiel Sprache, Handschrift, anthropomorphe Schnittstellen.
- Adaption von mobilen Umgebungen, so z.B. Laptops, Organizern, mobile Telefone.
- Einbindung von mobilen Assistenzfunktionen mit Bezug zum Aufenthaltsort des Benutzers (Location Awareness)

Die entsprechenden Arbeiten werden u.a. durch die offizielle Arbeitsstruktur (working structure) des Projektes MAP wiedergegeben, siehe Abbildung 6.5.

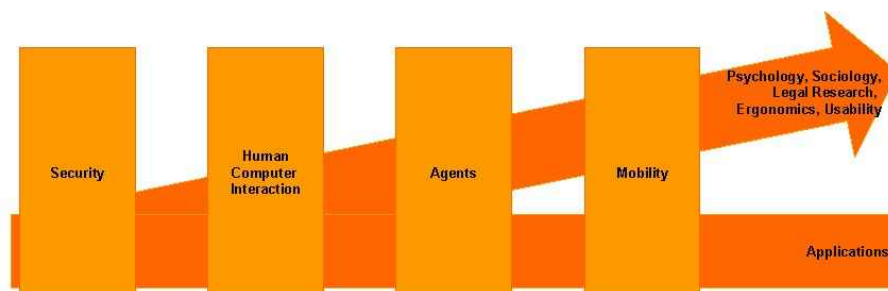


Fig. 6.5: Working Structure des MAP Projekts

Komplexe Koordination und adaptive Informationspresentation in Kombination mit mobiler und multipler Aufgabenbewältigung sind durch das MAP realisiert.

6.2.2 MAP Agentensystem

Das MAP System arbeite agenten-basiert. Dies bedeutet, dass MAP sowohl gewisse Basisagenten als auch eine Umgebung für Agenten zur Verfügung stellt.

Als Agenten werden im Projekt MAP Software Programme bezeichnet, die sich zumindest durch die folgenden Eigenschaften auszeichnen, vgl. [WJ95]:

- Autonomie: Ein Agent ist in gewissen Grenzen zu selbstständigem Handeln befähigt.

- Proaktivität: Um gewisse Zielvorgaben zu erreichen versucht der Agent aktiv an den Zielen zu arbeiten.
- Reaktivität: Der Agent kennt seine Umgebung und kann in gewissen Grenzen darauf reagieren.
- Kommunikativität: Ein Agent kann mit anderen Agenten kommunizieren um seine Ziele zu erreichen.

Um mehreren Agenten eine gemeinsame Basis zu bieten, welche den Agenten das Ausüben ihrer Agententätigkeiten ermöglicht, wird eine Multi-Agenten Plattform benötigt. Eine solche Plattform sollte zumindest eine verteilte Datenhaltung, Migrationsmöglichkeiten und Kommunikationsmöglichkeit anbieten.

Als MAP-Basissystem, vgl. [BSR⁺01], dient die Agentenplattform SeMoA (Secure Mobile Agents), vgl. [RP01], als Kommunikationssprache zwischen den Agenten dient FIPA-ACL, vgl. [ACL01]. Dieses Basissystem ist in der Lage, sowohl die Prozessrechnung als auch den entsprechenden Prozess-Code (die Programm-Ausführung als auch die Agenten selbst) dynamisch zu verteilen. Dies ermöglicht entsprechenden mobilen Agenten selbstständig im Netz zu agieren und zu transferieren, Informationen zu suchen und Aufgaben auszuführen - dies ohne eine ständige Verbindung zum Aufgaben-initialisierenden Benutzer aufrecht zu erhalten und damit ohne die Notwendigkeit einer ständigen Netzverbindung. Wenn ein Agent seine Ziele erreicht hat, transferiert er sich selbstständig zu seinem Benutzer und präsentiert seine Ergebnisse. Da ein mobiler Agent zwar Dienste anbietet und ein Spezialist auf seinem eigenen Gebiet ist, jedoch nicht zwingend ein höheres Benutzerinterface implementieren sollte, geschieht die Präsentation der Ergebnisse über einen speziellen Agenten - den User Interface Agent (UIA). Zu diesem Agenten existiert eine genau definierte Schnittstelle, welche die Übertragung von Ergebnissen zur Darstellung durch den UIA sowie die Rückübertragung von diesbezüglichen Interaktionen durch den Benutzer ermöglicht.

6.2.3 Der User Interface Agent

Der User Interface Agent ist die generelle Mensch-Maschine Schnittstelle des MAP Systems. Dies bedeutet insbesondere, dass der UIA eine Präsentations- und Interaktionsmöglichkeit zur Verfügung der Agenten des Agentensystems stellt. Diese Präsentations- und Interaktionsmöglichkeit muss zumindest folgenden Anforderungen genügen:

- Mensch-ähnliche, intuitive Kommunikation mit dem Agenten.
- Intuitive Repräsentanz verschiedener Agenten und damit verschiedener Zielvorgaben des Benutzers.
- Multimodale Ein- und Ausgabemöglichkeiten für Benutzer und für Agenten.
- Repräsentanz des individuellen, adaptierten MAP über diverse mobile Geräte.

- Umgebungs-Bewußtheit zur Ansteuerung geeigneter Medienkanäle.
- Pro-Aktivität.
- Zumindest eine anthropomorphe Ausprägung des Repräsentanten.



Fig. 6.6: Komponenten des User Interface Agenten und ihr Zusammenwirken, vgl. [KP01]

Die multimodale Ein- und Ausgabemöglichkeit für Benutzer und Agenten repräsentiert den medialen, Geräte-bezogenen Anteil des UIA. Dieser Anteil muss mit einer geeigneten Interaktionsmetapher verbunden werden, um tatsächlich die geforderte Mensch-ähnliche, intuitive Kommunikation mit den Agenten zu verwirklichen. Bild 6.6 zeigt eine Übersicht über die Komponenten und Datenströme des UIA. Die Architektur des UIA wird treffend durch das Bild 6.7 wiedergegeben.

Die höheren Funktionen des UIA wie Dialoginterpretation, Dialogmanagement und medienbeziehungsweise modalitätsbezogene Integration im Zusammenspiel mit der medialen Umsetzung via abstrakter Aktoren und deren konkreter Ausgabe ist in Bild 6.8 skizziert.

Als abstrakter Akteur wird dabei ein sinnzusammenhängender, für den Benutzer als solcher wieder erkennbarer MAP-Repräsentant gesehen, der dem Benutzer vertraute, menschähnliche, intuitive Interaktionsmöglichkeiten zur Verfügung stellt.

Durch diese Besonderheiten des MAP, die sich direkt aus dem Konversationsmodell des UIA ergeben, wird eine kombinierte sowohl implizite als auch explizite virtuelle körperliche Repräsentanz des Systems geschaffen: Ist der anthropomorphe Gesprächspartner nicht sichtbar oder hörbar, so ist er doch implizit im Dialog für den Benutzer zu erkennen. Hierzu werden psychologische Modelle menschlichen Verhaltens bemüht, die es erlauben, Art und Weise der Kommunikation auf dem Publikum bekannten Eigenheiten der konversationalen Schnittstelle des Systems MAP zu projizieren.

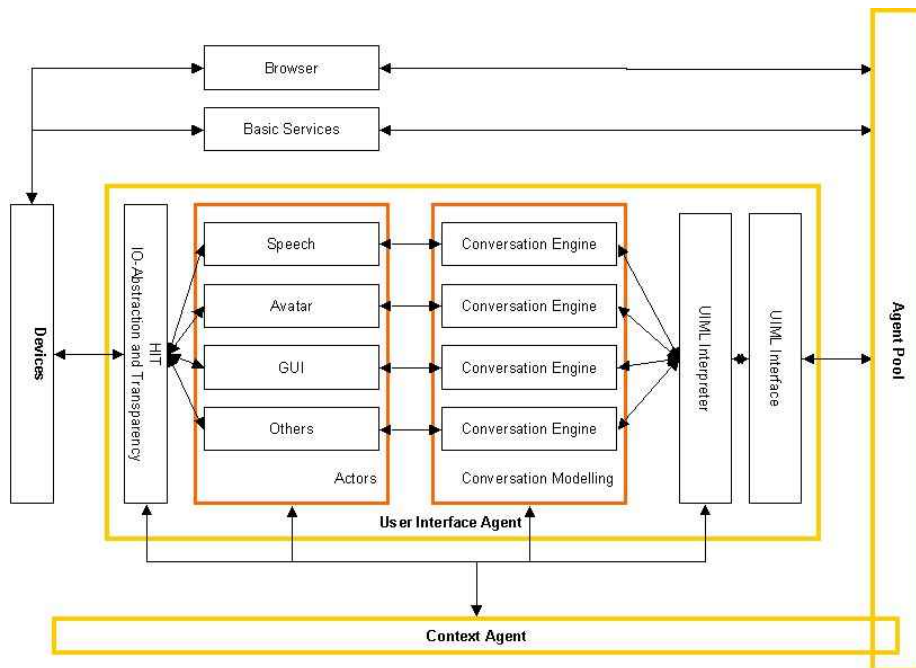


Fig. 6.7: Architektur des User Interface Agenten

Im MAP Projekt ist ein anthropomorpher Charakter ebenso realisiert wie ein akustischer Charakter oder eine rein graphische Bedienoberfläche. Die gewählte Interaktionsmetapher ist durch intensive Studien der Mensch-Mensch Kommunikation und Interaktion impliziert: Der User Interface Agent implementiert eine konversationale Interaktion mit dem Benutzer, die darauf ausgerichtet ist, jedwede Inhaltsinformation dem Benutzer in geeigneter, konversationaler Art interaktiv zu präsentieren.

Ein Beispiel soll den vorherigen Absatz weiter verdeutlichen: der Benutzer wendet sich an den User-Interface-Agenten mit einem Problem aus seiner Arbeitsumwelt - zum Beispiel einer Geschäftsreise. Warum wendet sich der Benutzer nicht direkt an die Agenten wie etwa einen Buchungsagenten für Hotel und Reise, einen Meetingagenten und einen Terminagenten? Der Grund hierfür ist, dass der Benutzer die Geschäftsreise als ein komplexes Ganzes⁶ sieht und die Geschäftsreise als ein ganzheitliches Problem bearbeiten möchte. Dies bedeutet insbesondere, dass sich der Benutzer nicht an verschiedene Agenten wenden möchte, um mit diesen den für die Agenten interessanten Teil der Geschäftsreise zu erarbeiten. Die Anforderungen an die Kommunikation mit dem Benutzer müssen aus der Sicht des Benutzers geschehen und nicht aus der Sicht des einzelnen Agenten. Der einzelne Agent mag wohl seine eigenen, speziellen Eingaben benötigen, für den Benutzer sind diese nur insoweit von Interesse, als er sein komplexes Problem wie die Geschäftsreise lösen kann. Genau für die Lösung von komplexen Problemen fehlt es den einzelnen Agenten jedoch an Überblick. Der Gesamtkontext muss daher in mindestens einer vermittelnden Einheit mit in Betracht

⁶ Vgl. 4.1.2.

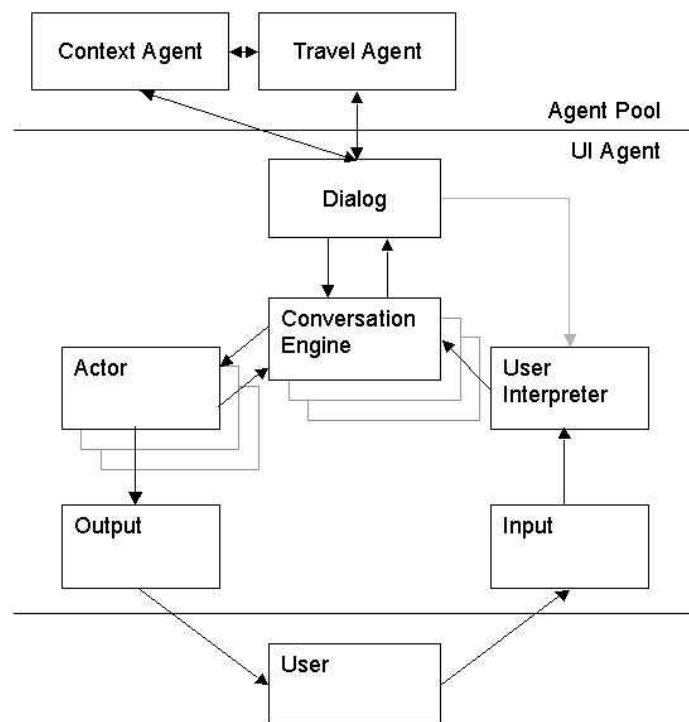


Fig. 6.8: Struktur des UIA bezüglich Dialog, Konversation und abstrakten Aktoren

gezogen werden, um die Kommunikation mit dem Benutzer im Interesse des Benutzers zu führen.

Des Weiteren wird ein kurzes Dialogbeispiel angeführt, was zur Illustration der oben genannten Anforderungen dient. Der Dialog beschäftigt sich mit einer Geschäftsreise. Es werden so unterschiedliche Gebiete wie Terminplanung, Meetingvereinbarung, Reisebuchung angesprochen - die einzelnen Posten werden nicht singulär abgearbeitet, sondern in einem Gesamtkontext zwischen Benutzer und System erörtert. Das Szenario⁷ dargestellt in Tabelle 6.5 skizziert diesen Aspekt des UIA.

Das Dialogszenario zeigt einige für Delegation und Assistenz typische Eigenschaften: der Benutzer drückt einen Zusammenhang aus, der in einem größeren Kontext betrachtet werden muss - pragmatisches Wissen über die Benutzersituation ist gefordert. Das System benötigt zum Lösen der Aufgabe mehrere Agenten. Diese sind im Beispiel die Agenten Terminkalender, Meetingagent, Flug-Buchungsagent, Bahn-Buchungsagent. Das System muss die Benutzerziele während der Konversation dynamisch anpassen - so ist im obigen Szenario eine Anpassung des Buchungsagenten notwendig.

⁷ Als Vorgabe für MAP stellt dieses Szenario die idealisierte Form der Kommunikation des Benutzers mit MAP dar - das für MAP notwendige Kontext-Wissen dieses Szenarios strebt gegen unendlich.

Tab. 6.5: Dialogszenario

1. Benutzer: Ich möchte morgen geschäftlich nach Berlin.
2. MAP: Geschäftlich!? Treffen Sie sich mit Projektpartnern?
3. Benutzer: Ja, mit Herrn Doktor Finke und seiner Gruppe.
4. MAP: Sie haben noch keinen festen Termin mit Herrn Dr. Finke - soll ich einen Termin für Sie vereinbaren.
5. Benutzer: Ja, um 14:00 Uhr wäre nett.
6. MAP: Darf ich einen Flug nach Berlin so buchen, dass sie morgen gegen 12:00 Uhr ankommen?
7. Benutzer: Bitte keinen Flug, ich fahre mit der Bahn.
8. MAP: Darf ich Ihnen eine Zugverbindung buchen, so dass sie morgen gegen 12:00 Uhr in Berlin ankommen.
9. Benutzer: Ja, mach das.
10. MAP: Ich möchte darauf hinweisen, dass die Zugverbindung morgen um 5:00 Uhr startet - dies überschneidet sich mit Ihrem Termin um 9:00 Uhr mit Ihrem Friseur.
11. Benutzer: Sag halt ab.
12. MAP: Ich sage Ihren Termin morgen mit Ihrem Friseur ab. Ich buche die Zugfahrt und vereinbare das Meeting mit Dr. Finke und seiner Gruppe.
13. Benutzer: OK, mach mal hin!

6.2.3.1 Agentenschnittstelle

Kontext- beziehungsweise Inhaltsinformationen werden in einem geeigneten UIML-Format von den Agenten an den UIA übergeben. Hierzu führt der UIA eine Schnittstelle, welche eben dieses Protokoll verstehen kann. Die UIML-Spezifikation der Schnittstelle ist definiert durch Blechschmitt, siehe [Ble01]. Blechschmitt definiert eine grundsätzliche, unabhängig von der medialen Ausprägung und Darstellung des Inhalts gehaltene Struktur. Diese ist in Tabelle 6.6 wiedergegeben.

Die entsprechenden Inhalte, wiedergegeben in UIML Dokumenten definiert aus Strukturelementen gemäß Tabelle 6.6 werden von einem entsprechenden UIML Interpreter interpretiert und anschließend in eine für eine Conversation Engine lesbare Art umgesetzt. Jedoch bildet dies nicht die gesamte Funktionalität des UIA ab - der Nutzer kann ebenso Ziele formulieren, für die erst noch die entsprechend zu delegierenden Agenten angesprochen werden

Tab. 6.6: Grundelemente des UIA - UIML Interface, [Ble01]

Markup Element	Funktion
action	Führt eine Aktion aus wenn die Bedingung der zugehörigen Regel erfüllt ist.
attribute	Eine Eigenschaft eines Part-Elements wird mit einem Wert versehen.
behavior	Ein Container für die Verhaltensdefinitionen.
call	Ruft eine Funktion eines Part-Element auf.
condition	Beinhaltet eine Bedingung.
constant	Beinhaltet eine Konstante.
content	Beinhaltet eine Menge an Konstanten.
uiml	Hiermit wird das UIML-Dokument als solches ausgewiesen.
equal	Prüft den Zustand eines (oder mehrerer) Ereignisses.
event	Beschreibt ein Ereignis, das zur Laufzeit erwartet wird.
head	Beinhaltet Metainformation über das Dokument.
interface	Beinhaltet die Dialogbeschreibung.
meta	Definiert ein Name/Wert-Paar für Metadaten.
param	Definiert einen Parameter z.B. zur Übergabe bei einem call-Element.
part	Definiert ein Dialog-Element.
property	Weist einer Eigenschaft den angegebenen Wert zu.
rule	Beinhaltet eine Regel bestehend aus Bedingung und Aktion.
structure	Beinhaltet die Spezifikation der Interfacestruktur.
style	Beinhaltet die Werte der Eigenschaften der Interfaceelemente.

müssen.

6.2.3.2 Dialoginterpreter

Was geschieht nun innerhalb des Systems, wenn der Benutzer den ersten Satz spricht? Der User-Interface-Agent verfügt über eine Dialog Interpreter, vgl. [BS01a]. Dieser überprüft den Satz auf verschiedene Schlagworte, so zum Beispiel Datum, Ortsangabe, Angabe von Gründen. Anschließend zitiert das System (über den Benutzerziel-Agenten) alle relevanten Agenten und überprüft, welche (eventuell gemeinsam zu nutzenden Daten) die Agenten benötigen, um eine sinnvolle Aufgabe zu erfüllen. Anhand der Dialog-Eigenschaften der Agenten wird eine entsprechender Baum von Dialogkomponenten aufgebaut. Hiernach wird das erste Element des Baums abgearbeitet, was in diesem Falle die Vermutung eine Grundes für die Reise ist. (Eine solche Vermutung bzw. eine Einschränkung des Kontextes ist sinnvoll, um die Konversation auf das wesentliche zu beschränken). So fragt das System also nach dem Grund der Reise, indem es ein Treffen mit Projektpartnern vermutet, siehe Punkt 2 des Szenarios 6.5. Nachdem der Benutzer die Vermutung bestätigt und mit der Bestätigung weitere Daten an das System übergibt werden weitere Agenten hinzugezogen.

Die Dialogkomponenten werden revidiert und der Baum neu aufgebaut. Die Frage nach dem Grund der Reise wird in dem neuen Baum nicht mehr berücksichtigt, da sie schon beantwortet ist. Dieser Vorgang zieht sich hin bis das Benutzerziel erreicht ist.

6.2.3.3 Konversationale Modellierung

Die Mengen von Dialogkomponenten werden vom UIML Interpreter an die entsprechende Conversation Engine übergeben. Deren konversationale Modellierung besteht im wesentlichen aus einer Menge von abstrakten Konversationsaspekten sowie Konversationsregeln⁸, die die Konversationsaspekte modifizieren und damit konversationale Situationen abbilden. Um die entsprechenden konversationalen Situationen zu bilden bzw. die funktionale Abbildung der konversationalen Situationen zu veranlassen, wird eine Schnittstelle zur Eingabe von Basis-Aspekten wie Diskursen (oder auch Diskurswechseln) sowie zur Eingabe von Inhaltselementen, die in Diskursen mittels Konversation dem Benutzer zugänglich gemacht werden, benötigt. Die Art und Weise, wie die Konversation von der Conversation Engine (CE) umgesetzt wird, kann über die Schnittstelle nicht beeinflusst werden. Die Conversation Engine bietet ein entsprechendes Protokoll an, dessen Grundelemente in Tabelle 6.7 beschrieben sind.

Tab. 6.7: Grundelemente des Conversation Engine Interface

Markup Element	Funktion
Diskurs	Bezeichnet einen Diskurs. Der Diskurs steht für eine Delegations-Situation bezüglich eines Agenten und strukturiert somit Dialogkomponenten.
Story	Ein Diskurs kann aus mehreren zusammenhängenden Inhalten, Fragen oder Antworten bestehen, welche in Storys gegliedert sind.
Inhalt	Ein Element, welches keinen direkten Bezug zu einer vorherigen oder nachfolgenden Dialogkomponente besitzt.
Antwort	Ein Element, welches einen direkten Bezug zu einer vorherigen Dialogkomponente besitzt.
Frage	Ein Element, welches einen direkten Bezug zu einer nachfolgenden Dialogkomponente besitzt.
Anfrage	Eine Struktur, welche zur Definition von Rückfragen der Conversation Engine bezüglich der Dialogkomponenten dient.

Die konversationale Modellierung innerhalb der jeweiligen Conversation Engine erfolgt mittels eines sogenannten Regel-basierten Systems, auch Expertensystem oder Wissensbasiertes System, vgl. [Ric98], genannt. In diesem System kann ein konversationaler Aspekt samt seiner Eigenschaften⁹ explizit als symbolische Wissenskomponente abgelegt werden.

⁸ Exakt jene Konversationsregeln, welche in Kapitel 5.2.2 definiert sind

⁹ Vgl. Kapitel 5.2.2.

Wissensbasierte Systeme sind aufgrund Ihrer Eigenschaften besonders geeignet, um Phänomene der Mensch-Mensch Kommunikation auf die Mensch-Maschine Kommunikation zu übertragen:

- Lösung intelligenter Phänomene auf der Ebene der Begrifflichkeit (nicht numerisch).
- Probleme werden Top-Down analysiert - dies in der Annahme, dass sich das rational-menschliche Denken auf symbolischer Ebene simulieren lässt, wenn sich entsprechende menschliche Begrifflichkeiten (Worte, sprachliche Gebilde) in Symbolik und Bedeutung finden lassen.
- Prozeduralität in Form von logischen Schlüssen lässt sich mit Hilfe von symbolverarbeitenden Regeln automatisieren.

Als Formalismus für die deklarative Beschreibung des Wissens verwendet die konversationale Modellierung des MAP Produktionsregeln. Eine Produktionsregel nutzt eine WENN - DANN Form der Informationsdarstellung: eine situationsbedingte Prämisse wird als Auslöser einer Aktion deklariert.

Mittels des sogenannten Modellierungsansatz, vgl. [Ric98], lässt sich Expertenwissen bezüglich der Mensch-Mensch Kommunikation, wie es zum Beispiel Designer, Soziologen oder Psychologen besitzen, in einem Prozess der Suche nach angemessenen Modellumrissen der konversationalen Modellierung, erwerben. Der Zweck dieser Modellbildung ist dabei nicht die Abbildung existierender Realität der Mensch-Mensch Kommunikation, sondern eine vereinfachte Konstruktion zur Übertragung auf die Mensch-Maschine Kommunikation. Die Modellbildung wird dabei als iterativer Prozeß gesehen, welcher in gewissen Stufen eine Feinabstimmung und Detaillierung beziehungsweise eine Revision ermöglicht.

Aus den oben genannten Gründen kann die symbolische Wissensrepräsentation durch einen Konversationsdesigner oder sonstigen User Interface Entwickler einfach manipuliert werden, da die Eigenschaften, die den Aspekt bestimmen, im Wesentlichen nicht-numerisch und menschenlesbar sind. Ebenso sind die konversationalen Regeln, welche die Aspekte manipulieren, lesbar, nicht prozedural, insbesondere Situations-bezogen ausgelöst, im System als Produktionsregeln auf den Aspekten gespeichert.

Zusätzlich wird durch den Verzicht auf eine explizite Prozeduralität und die Übergabe der Programmausführung an eine logische Maschine mit einem entsprechenden logischen Kalkül eine deklarative Art der Programmierung unterstützt, die vor allem den Autoren und Designer von konversationalen Diskursen entlastet.

6.2.4 Konzeptumsetzung

Im MAP Projekt kann nahezu das gesamte Konzept, wie es in Kapitel 5 definiert ist, umgesetzt werden. Die Teilung des Konzeptes in separate Bereiche für die konversationale Modellierung, die medienspezifische, direkt manipulative Interaktion und die Einbindung von Content-Provider Applikationen als Kontext-gebende Einheiten, welche über ein gemeinsames User Interface mit dem Benutzer kommunizieren, entspricht den Anforderungen an ein

System zur interaktiven Narration von kontinuierlichen Inhalten. Lediglich die tatsächliche narrative Komponente in den inhaltsgebenden Applikationen ist nicht nach dem Konzept des Kapitels 5 umgesetzt.

Die konversationale Modellierung im System geschieht auf abstrakte, explizite, amodale Art im User Interface Agenten. Agenten-Applikationen schicken ausschließlich Kontent-bezogene Informationen an die Conversation Engine, die diese Informationen selbstständig um konversationales Verhalten erweitert und an entsprechende Aktoren gibt. Die Aktoren setzen die Konversation modal um. Die Benutzereingabe-interpretierende Komponente wiederum bildet die modale Eingabe des Benutzers auf die abstrakten Konversations- bzw. Inhaltselemente ab, wodurch sie durch die Conversation Engine und im Anschluss durch die Applikationen ausgewertet werden können.

Die direkt manipulative Interaktion ist medienspezifisch durch die Applikationen selbst gelöst, was eine optimierte Interaktion zulässt ohne den Umweg über eine konversationale Schnittstelle. Die Narration der Informationen geschieht weitgehend unabhängig vom MAP System, jeder Agent hat die Möglichkeit selbstständig seine Informationen zu strukturieren. So existiert im MAP System zum Beispiel ein Videokonferenzen Agent welcher eine direkt manipulative Annotation von Video Live-Streams erlaubt. Diese geschieht jedoch intermedial, d.h. die Annotation erfolgt nicht innerhalb des Videostreams sondern als Text mit einer zeitlichen Zuordnung zum Video. Eine explizite Story Modellierung ist im MAP System nicht vorgesehen.

6.3 DIVA

6.3.1 Einleitung

Das System DIVA - Digital Interactive Video and Audio - stellt eine interaktive Video- und Audio-Präsentationsumgebung dar. Alle Interaktionsmöglichkeiten, die das DIVA System anbietet, werden in den Kontext der präsentierten Video-Clips gesetzt - quasi vom Video als solchem angetrieben - und damit über ein medienzentriertes User Interfaces bearbeitet. Interaktionsmöglichkeiten werden dem Publikum der Videopräsentation sowohl Konversational (bezogen auf die Inhalte des Videos) oder durch (klassische) direkt manipulative Interaktion (auf die sicht- und hörbaren Objekte im Video) angeboten. Die Möglichkeiten der Konversation sind dabei sowohl anthropomorph - über einen menschähnlichen Ansprechpartner beziehungsweise Assistenten - oder direkt auf das Video gegeben.

Mit dem DIVA System wird gezeigt, dass eine konversationale Interaktionsmöglichkeit bei medienzentrierten User Interfaces, vgl. Braun [Bra01a], tatsächlich für einen Nutzer zu handhaben ist. Unter einem medienzentrierten User Interface wird dabei ein Interface zum Benutzer verstanden, das die Interaktion mit dem Medium als solches - im Falle von DIVA ist das Medium Video gewählt - unterstützt und die benötigten Interaktionsformen zur Verfügung stellt, um die Informationsverarbeitung durch den Benutzer intuitiv und medienbezogen zu gestalten.

Durch die Kombination von direkt manipulativen und konversationalen Interaktionsmög-

lichkeiten sollen die temporal bedingten Konflikte der Videopräsentation überwunden werden¹⁰:

Der Ansatz des DIVA Systems der kombinierten direkt-manipulativen und konversationalen Interaktionsmöglichkeit auf ein kombiniert selektiv/generatives videobasiertes Storytelling-System nutzt diverse State-of-the-Art Technologien wie Sprachausgabe, Spracheingabe, Videoservice und Animation, um das Konzept zu realisieren. Bezüglich folgender Punkte wurden eigene Ansätze verfolgt:

- Ein Konversationsansatz, basierend auf einem menschähnlichen Avatar.
- Ein direkt manipulativer Ansatz, basierend auf Video-Annotation.
- Nichtlineare Story-Narration basierend auf einem selektiven und generativen Ansatz.

Die Implementierung des Ansatzes wird anhand einer Real Media Server, SMIL und Java 3D basierten Applikation gezeigt. In den folgenden Unterkapiteln wird die Umsetzung des Konzeptes erläutert, in dem sowohl der nichtlineare, kontinuierliche Medienservice, die konversationale Interaktionskomponente als auch das Zusammenspiel der Komponenten diskutiert werden.

6.3.2 Story & Videoservice

Der Videoservice muss verschiedene Funktionalitäten bieten, um sowohl die Video-Clips als auch die Annotationen, deren zugehörige Annotationsziele wie auch die Konversationsdaten übertragen zu können. Weiterhin muss er eine Interaktionsmöglichkeit auf den Annotationen bereitstellen sowie entsprechende Schnittstellen zu den Applikationen, die die konversationale Interaktionsmöglichkeit bereit stellen. Der resultierende Datenfluss über den Videoserver ist in Abbildung 6.9 dargestellt.

Video Service ist in verschiedenen Varianten bereits als Komplettsystem erhältlich. MPEG4, vgl. [Koe01], stellt die zur Zeit aktuellste Entwicklung dar/footnoteVgl. Kapitel 3.3.1.1., verfügt aber z.Z., vgl. [Koe99], nicht über einen interaktionsfähigen Video Player.

Der sogenannte Media Player, vgl. [Obe99], von Microsoft¹¹ kann Multimediaclips nur nacheinander und nicht parallel abspielen. Aus diesem Grund fällt er bezüglich des parallelen Abspielens von akustischen Annotationen aus.

Der Videoserver von Real¹² verfügt über die Funktionalitäten, die zum Erzeugen einer nichtlinearen Videopräsentation mit zusätzlichen, parallel übertragenen Informationen ideal¹³ sind, vgl. [Gra00].

¹⁰ Vgl. Kapitel 4.3.2.

¹¹ Vgl. Kapitel 3.3.1.1.

¹² Vgl. Kapitel 3.3.1.1.

¹³ Durch die Möglichkeit, sowohl serverseitig als auch clientseitig entsprechende Codecs einzusetzen, welche u.a. textuelle Informationen als Datenstrom (Stream) versenden, kann sowohl eine parallele (synchrone) Kontribution von zusätzlichen Daten neben dem Video geboten als auch eine asynchrone Datenübertragung realisiert werden, vgl. 3.3.1.1.

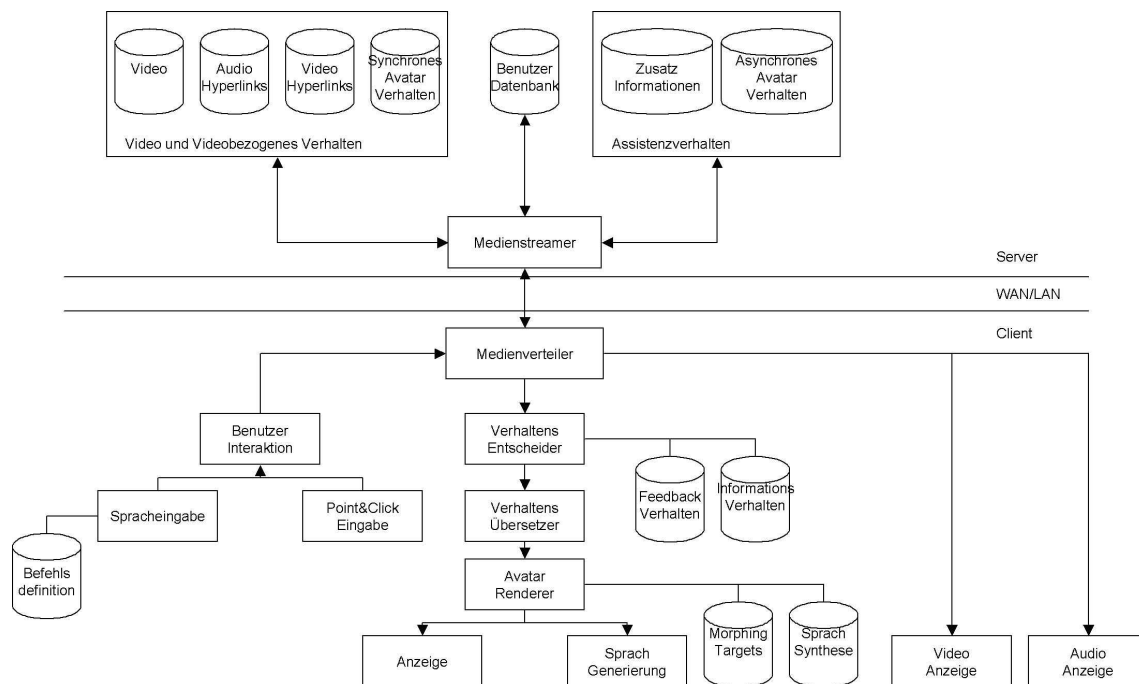


Fig. 6.9: Datenfluss im DIVA Systems

6.3.2.1 Umsetzung von Hyperlinks

In Bezug auf das Medium Video stellt die Kontinuität des Mediums eine besondere Herausforderung im Hinblick auf direkt manipulative Ansätze dar¹⁴. Diese Ansätze werden gemäß des Konzeptes in Kapitel 5 umgesetzt, d.h. als temporal hypermediale, intramediale Annotation des Mediums.

Die Darstellung der akustischen und graphischen Hyperlinks erfolgt als temporaler, expliziter Hyperlink¹⁵.

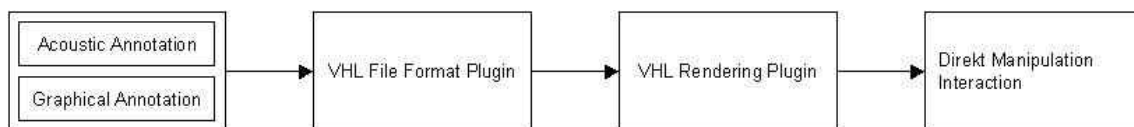


Fig. 6.10: Applikationsspezifische Speicherung von graphischer und akustischer Annotationsinformation

Um eine graphische Annotation darzustellen und dem Benutzer visuell die Position und die Verweildauer eines Hyperlinks anzeigen zu können, bedarf es zweier geeigneter Plugins, siehe Abbildung 6.10:

¹⁴ Vgl. Kapitel 4.3.2.

¹⁵ Vgl. Kapitel 4.3.2.1 und Kapitel 3.3.1.1.

- Ein File-Format-Plugin, VHL File Format Plugin genannt, zum Lesen der Darstellungsinformationen und Bereitstellung der Datenpakete zum kontinuierlichen übertragen (dem sogenannten Streamen) von einer Datenquelle zum Datenziel.
- Ein Rendering-Plugin, fortan VHL Rendering Plugin genannt, zur Darstellung auf dem Bildschirm.

Da ein Videohyperlink seine Größe und die Position innerhalb eines Videos während seiner Gültigkeitsdauer verändern kann¹⁶, muss diese Information in geeigneter Art und Weise übertragen werden. Dabei sind zwei Ansätze möglich:

Client-lastig Bei der Client-lastigen Variante werden dem VHL Rendering Plugin Startposition, Startzeitpunkt, Endposition, Endzeitpunkt und die Form (Kreis, Rechteck, Polygonzug) übermittelt. Dazu erhält es noch eine komplexe Zusatzinformation, die den Verlauf der Bewegung beschreibt und die Größe des Objekts definiert.

Authoringtool-lastig Bei der Authoringtool-lastigen Variante erhält das VHL Rendering Plugin Position, Startzeitpunkt, Endzeitpunkt, Größe und Form (Kreis, Rechteck, Polygonzug) eines Hyperlinkabschnittes. Damit das Rendering Plugin weiß, wie viel ein solcher Hyperlinkabschnitt vom Gesamthyperlink darstellt, erhält es zusätzlich zwei Prozentangaben: eine, die angibt, wie viel Prozent des gesamten Hyperlinks schon gezeichnet wurden, und eine, die angibt, wie viel Prozent des gesamten Hyperlinks dieser Abschnitt darstellt.

Die Client-lastige Variante ermöglicht zwar ein sehr einfaches Authoringtool, allerdings wird das Rendering Plugin sehr komplex: eine Bewegung kann nun einmal die unterschiedlichsten Formen haben. Soll z.B. eine Person über Hyperlinks anwählbar sein und läuft diese Person während einer Szene im Bild von links nach rechts und wieder nach links, so werden hierzu Zeit, Position und Verweildauer der einzelnen Abschnitte benötigt. Das Authoring-Tool müsste auch hier die Bewegung in einzelne Stücke unterteilen und diese in das passende Format für die Bewegungsabläufe abbilden.

Die Authoringtool-lastige Variante hat den großen Vorteil, dass das Rendering Plugin nur wenig Komplexität beinhaltet und sich trotzdem selbst komplexeste Bewegungsabläufe darstellen lassen - wenn das Authoringtool dies ermöglicht. Nachteilig ist lediglich, dass das Hyperlinkziel (die mit der Annotation verbundene URL) bei bewegten Hyperlinks mehrfach übertragen wird. Da dies aber nur bei komplexen Hyperlinks der Fall ist, welche vermutlich nur selten Anwendung finden, wird die Authoringtool-lastige Variante für das System gewählt.

Die Speicherung der Annotationsdaten ist als Anwendungsspezifisch identifiziert, jedoch zeigt sich, dass eine deklarative Art des Ablegens von Annotationsinformation eine einfache und leistungsfähige Möglichkeit des Managements entsprechender Informationen ist.

¹⁶ Vgl. Kapitel 3.3.1.1.

Das Synchronized Multimedia Integration Language Protokoll¹⁷ (SMIL) erlaubt dies in umfassender Art und Weise.

Die akustische Annotation wird in einer Audio-Datei abgelegt und ist mit den vorhandenen Möglichkeiten von RealSystem abgedeckt. Um die akustische Annotation nicht fest mit dem Video zu verbinden, werden die Annotationstöne über SMIL parallel zum Video abgespielt: Entweder als Datei in der Länge des Videos, die alle Annotationen enthält oder indem ein Annotationston (oder auch mehrere unterschiedliche) über SMIL mehrfach abgespielt wird.

6.3.2.2 Nichtlineare Story Narration

Die nichtlineare Erzählung der Geschichte basiert im wesentlichen auf dem Zusammenfügen von Video-Clips. Diese Video-Clips werden mit entsprechender Metainformation versehen, welche Ihre Auswahl aufgrund von abstrakten Inhalts- oder Funktionsinformationen gewährt. Sowohl eine branching-basierte Version als auch eine Version basierend auf morphologischen Storyfunktionen¹⁸, wurden getestet.

- Branching-Ansatz: Durch den Autor der nichtlinearen Präsentation sind vorgegebene hypermediale Bezüge zwischen den Video-Clips als Basis der nichtlinearen Narration erstellt. Weiterhin wird eine Zuordnung von inhaltlicher Information (durch eine auf einer Sprachgrammatik und Wortanalyse basierenden Interpretation) zu Video-Clips vorgenommen.
- Morphologischer Ansatz: Eine funktionale Bewertung der Videos wird als Grundlage der Zusammenstellung der Video-Clips verwendet. Als Auswahlkriterium einer Story Engine dienen dabei Story Acts¹⁹, welche durch die Interaktion des Benutzers auf dem Video ausgelöst werden. Die Möglichkeiten der Zuordnung des Story Acts hängt dabei davon ab, wie die Rolle des Avatars in der Geschichte²⁰ definiert ist. Kann eine sinnvolle Interaktionsmöglichkeit geboten werden, die aus zumindest zwei Story Acts auswählt, so ist ein nichtlinearer Aufbau auf morphologischer Basis sinnvoll. Ansonsten wird im wesentlichen auf Seitenfaktoren wie Altersfreigaben, Zeitvorgaben und bis zum jeweiligen Zeitpunkt gezeigte Clips eingegangen.

Das in Kapitel 5.3.1 vorgestellte Konzept der Story Engine auf Basis von Propp dient dabei als Ausgangspunkt der nichtlinearen Narration. Durch das Hinzufügen des Konzeptes der *Moves* kann eine konsistente Umgebung bezüglich der Darsteller etc. definiert werden.

¹⁷ Vgl. Kapitel 3.3.1.1.

¹⁸ Vgl. Kapitel 5.3.1.

¹⁹ Mögliche Story Acts sind zum Beispiel das Verhindern der Propp-Funktion *Anschlag des Feindes*. Durch Verhindern des Anschlags müssen entsprechende morphologische Funktionen als Ersatz ausgewählt und präsentiert werden.

²⁰ Vgl. Tabelle 5.2.

6.3.3 Konversationsaspekt

Die drei Möglichen Einsatzformen des Avatars²¹, werden alle durch die vorgegebenen Kontexte - die Inhalte des präsentierten Videos - gesteuert. Die Nutzung ist somit zentriert auf das darzustellende Medium. Als besonderer Benefit wirkt dabei die Reduzierung des konversationalen Kontexts durch die Video-Clip Inhalte. Erst durch diese Reduzierung ist die sinnvolle, automatische Gestaltung eines Gesprächs möglich und kann in Form einer Konversationsmodellierung erfolgen.

Tab. 6.8: Konversationale Einsatzweise eines Avatars zum Video

Einsatz	Beschreibung
Synchron zum Video	Während das Video präsentiert wird, gibt der Avatar synchron zum Videoablauf definierte Inhalte und Verhaltensweisen wieder. Die Art der Verhaltensweisen und Informationen sind abhängig von der Art des narrativen Einsatzes des Avatars, siehe Tabelle 5.2.
Asynchron zum Video	Über die konversationale Interaktion wird ein indirekter Zugriff auf die Inhalte des Videos geboten. Dies ermöglicht eine asynchrone, Benutzer-angepasste Darbietungsmöglichkeit von Inhalten und Verhalten - unabhängig davon, ob die Inhalte des Videos gerade sichtbar sind oder nicht.
System-Feedback zum Benutzer	<p>Jede Interaktion des Benutzers bedarf eines direkten Feedbacks durch das System - dieses Feedback ist zwar in Abhängigkeit von der synchronen beziehungsweise asynchronen Interaktionsmöglichkeit des Benutzers zu sehen, stellt aber keine inhaltliche Antwort des Systems dar, sondern eine diskurs-bezogene Antwort. Jede Interaktion des Benutzers benötigt ein definiertes Feedback des Systems [Bau95]. Das Feedback kann von folgendem Typ sein:</p> <ul style="list-style-type: none"> • Assistenz-Feedback wenn der Benutzer direkt mit dem Video interagiert hat. • Navigationales Feedback bezüglich der Video Abspieler (VCR) Funktionalität. <p>Das Feedback wird durch einen Avatar gegeben, der dadurch durch das ganze System führt.</p>

Die konversationale Interaktionsmöglichkeit mittels eines Avatars auf dem Kontext des Videos²², die in dem Video erfahrbar sind - werden nicht durch zeitliche Limitationen

²¹ Vgl. Kapitel 5.3.3.

²² D.h. bezüglich der Objekte im Video, welche konkret (z.B. ein Haus, das in einem Video zu sehen ist) oder abstrakt (z.B. der Name eines Schauspielers oder der Konflikt zwischen Spielfiguren) sein können.

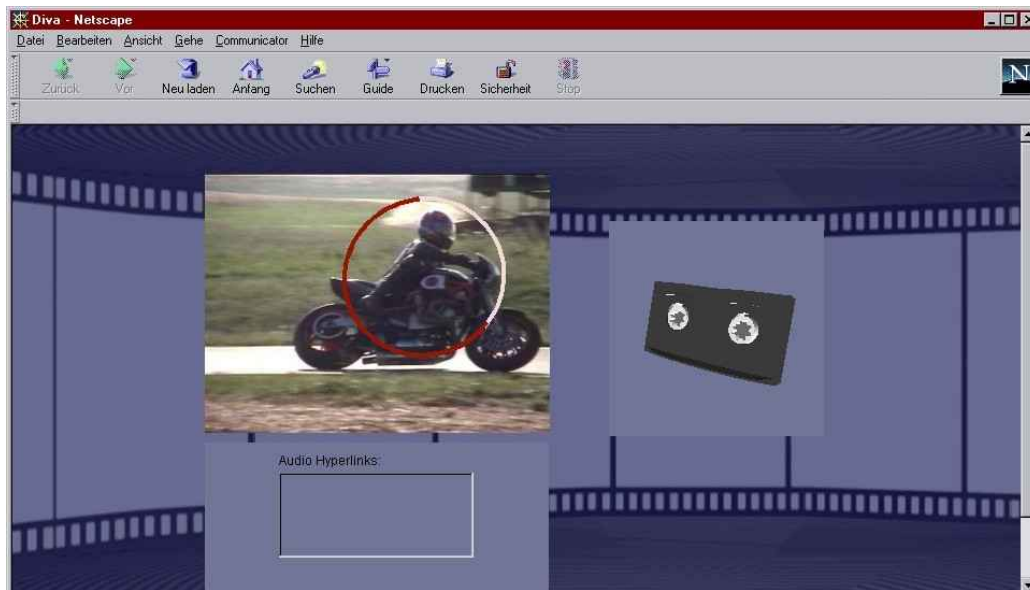


Fig. 6.11: Konversation mit menschähnlichem Avatar, angetrieben durch den Video Kontext (Screenshot des DIVA Systems)

der Sichtbar- oder Hörbarkeit von Objekten eingeschränkt, siehe Bild 6.11. Aus diesem Grund kann die konversationale Interaktionsmöglichkeit sowohl für synchrone als auch für asynchrone Interaktionen auf den Videoinhalten genutzt werden. Werden beide Interaktionsmöglichkeiten genutzt, so reduziert dies die Konflikte, die sich aus den temporalen Eigenschaft des Videos ergeben. Die Konversation bzw. Inhaltsdarbietung durch den Avatar folgt dabei den Prinzipien, dargestellt in Tabelle 6.8.

6.3.3.1 Charakter

Die Rolle des Ansprechpartners im System übernimmt ein Avatar. Die Persönlichkeit und das Aussehen des Avatars ist grundsätzlich je nach Informationstiefe veränderbar. Das bedeutet, dass bei Information über ein bestimmtes Genre bzw. bei der Präsentation eines Videos zur normalen Darstellung des Avatars ein genrespezifisches Merkmal hinzugefügt wird. In allen anderen Fällen ist er - im Hinblick auf die große Zielgruppe im Web - neutral modelliert, jedoch besitzt er dennoch einen hohen Wiedererkennungswert.

Als Charakter des Avatars sind zwei Varianten in DIVA vorhanden. So kann einmal ein menschähnlicher Charakter eingesetzt werden. Als Charakter kann auch eine Applikations- und damit Video-spezifische Figur eingesetzt werden, die zwar anthropomorphe Formen aufweist, allerdings sofort als nichtmenschlich zu erkennen ist. Im Falle von DIVA ist dies eine animierte Video Kassette, siehe die Abbildungen 6.12, 6.13 und 6.14.

Das Avatar Verhalten wird nicht als Animation vorab gespeichert, sondern tatsächlich durch das Ansteuern eines Java3D Feature Morphing Renderers, vgl. [AM99], explizit zur Laufzeit

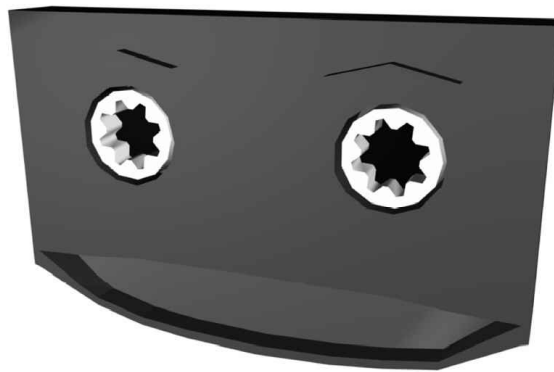


Fig. 6.12: Videospezifische Darstellung des Avatars

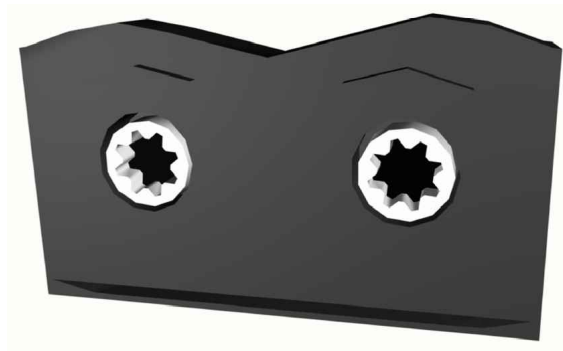


Fig. 6.13: Videospezifische Darstellung des Avatars, Augenbraue hochgezogen

des Systems erzeugt. Dieser Renderer realisiert die in Kapitel 5.2.2.2 notierte Funktionalität. Die Verhaltensübertragung geschieht mittels einer XML Struktur.

6.3.3.2 Verhaltensentscheider

Der Verhaltensentscheider ist ein Regelsystem, dass das aktuelle Verhalten auf Basis der konversationalen Benutzereingaben und des aktuell zu zeigenden, Präsentations-spezifischen Verhaltens definiert. Der Charakter kann direkt vom Verhaltensentscheider²³ angesprochen werden. Das Verhalten setzt sich aus drei Komponenten zusammen:

- Synchron: Das Verhalten, das synchron zur Video Präsentation an den Client verteilt wird.
- Asynchron: Asynchrone, konversationale Verhaltensweisen des Avatars, die durch die konversationale Modellierung erfolgen.
- Feedback: Auf dem Client generiertes Feedbackverhalten.

²³ Vgl. Kapitel 6.3.3.2.

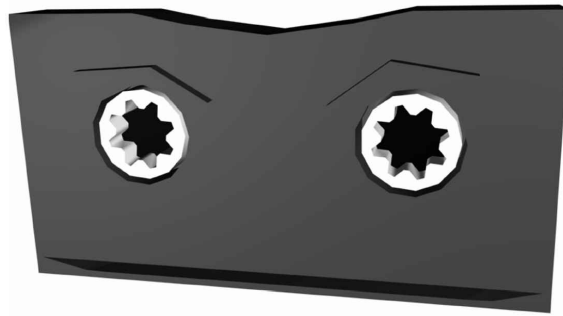


Fig. 6.14: Videospezifische Darstellung des Avatars, Emotion *Veraergert*

Laut den Prioritätsvorgaben des Autors der Präsentation wird darüber entschieden, welches Verhalten der drei Verhaltenskomponenten den Vorrang hat. Aufgrund dieser Entscheidung wird die Beschreibung des entsprechenden Verhaltens an den Avatar-Renderer übermittelt.

Die Entscheidung über das Avatar-Verhalten erfolgt nach folgenden Überlegungen:

- Auf jede Benutzereingabe wird sofort reagiert, da sie den aktuellsten Wunsch des Benutzers wiedergibt. Dieses Verhalten gibt dem Benutzer ein Feedback, das als Turntaking-Akzeptanz aufgefasst wird.
- Liegt eine konversationale Eingabe des Benutzers vor, so wird mit der entsprechenden konversationalen Ausgabe des Avatars sofort darauf reagiert.
- Liegt keine Benutzereingabe vor, so wird das synchrone Avatarverhalten zum Video gezeigt.
- Der Avatar darf nie aufhören sich zu verhalten. Dies wäre zum Beispiel der Fall, wenn der Benutzer ein Video stoppt und danach nicht gleich einen weiteren Request durchführt. Für solche Situationen ist ein Füllverhalten zu definieren.
- Ebenso wird für die Zeit zwischen einem Request und der Systemantwort ein Warteverhalten festgelegt.

6.3.4 Informationshaltung & Service

Im Folgenden werden die für den Service der nichtlinearen Video-basierten Geschichte notwendigen Server-, Middleware- und Clientkomponenten beschrieben, wie sie in Abbildung 6.15 zu sehen sind.

Server Das System nutzt einen Real Video Server und eine Datenbank um die selektiven und generativen Elemente einer Geschichte sowie Meta-Informationen abzulegen. Abbildung 6.15 zeigt sowohl einen Pool von synchronen, asynchronen als auch

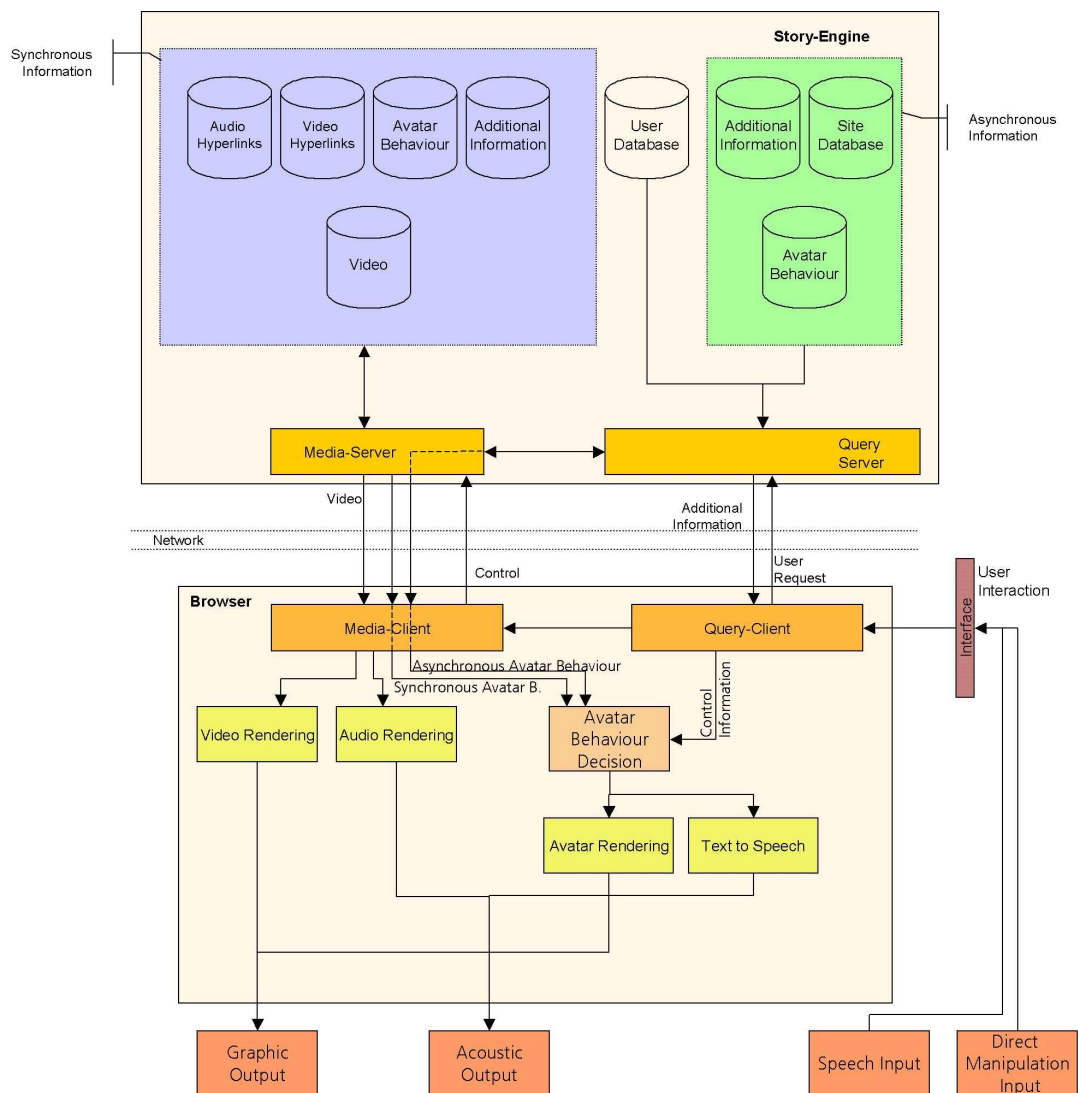


Fig. 6.15: Komponenten des DIVA Systems

Benutzer-spezifischen Informationen, die durch das System vorgehalten werden. Diese Informationen bestehen aus Videos, der hypermedialen Annotation (temporal, spatial, hypermediale Ziele), zusätzlicher Information zur Konversationsführung in textueller Form, synchrones Avatarverhalten sowie (SMIL-) Synchronisationsdaten. Diese Informationen werden genutzt, um abhängig von der Benutzereingabe und den allgemeinen Vorgaben einer Story Engine die zu zeigenden Video-Clips und Verhaltensweisen des Avatars zu selektieren bzw. zu generieren. Dieses sind die grundsätzlichen, konversationalen Verhaltensweisen des Avatars sowie das Feedback-Verhalten des Avatars. Neben den direkt Story-bezogenen Daten werden zusätzlich Kontroll- und Verwaltungsdaten, zugänglich über den Query-Server, bezüglich der Benutzer abgelegt, um

das System einerseits Mehr-Benutzer fähig und andererseits mit einer gewissen Datensicherheit auszustatten.

Middleware Die Story Elemente werden mittels des Real Media Servers an einen Real Media Client geschickt. Die Multimediasynchronisation geschieht mittels SMIL. Tatsächlich erfolgt die Übermittlung von Video, Annotationen und Avatar-Verhalten in separaten Streams, welche durch SMIL synchronisiert werden, siehe Abbildung 6.15. Zusätzlich zum unidirektionalen Versenden von Informationen ist ein bidirektionaler Kanal zwischen den Anfrage-Komponenten des Systems (Query-Server und Query-Client) realisiert, der die konversationalen Elemente der Interaktion zwischen Benutzer und Video überträgt.

Client Der Client ist geteilt in einen Media Client und in eine Query Client Komponente.

Der Real Media Client verteilt die Daten - Video und Avatar Verhalten inklusive Avatar-Text - zu folgenden Komponenten:

- Real Video Renderer
- Avatar Verhaltensentscheider
 - Avatar Renderer: Java-3D basierter Feature-Morphing Renderer.
 - Viseme-Generator
 - Akustische Ausgabe: Phonemgenerator, vgl. [PSP⁺92], in Kombination mit einem Sprachgenerator, vgl. [DPP⁺96].

Das Avatar Verhalten, kombiniert mit einer Lippensynchronisation, wird auf dem Client mittels des Verhaltensentscheiders generiert.

Die konversationale Interaktion des Benutzers wird mittels einer Spracherkennung vorverarbeitet, dann mittels eines Konversationsmodells zu symbolischen Anfragen über den Query-Client an die Server Datenbank verarbeitet. Antworten des Servers werden über den Real Media Client an den Avatar gegeben.

Die Sprachverarbeitung erkennt zusätzlich die akustische Interaktion auf die akustischen Teile des Videos und gibt diese als entsprechende Befehle an den Real Media Server weiter. Die graphischen Annotationen des Videos werden von dem Real Media Client selbst verarbeitet und an den Real Media Server geschickt, wo sie unter anderem der Story Engine als Eingabe dienen.

Die Komponenten ermöglichen einen Service der nichtlinearen Geschichte über das World Wide Web sowie eine Darstellung der Story in einem entsprechenden Web-Browser.

6.3.5 Konzeptumsetzung

Die Umsetzung des in Kapitel 5 vorgestellten Konzepts gelingt mit dem DIVA System fast vollständig. DIVA enthält folgende Komponenten:

- Eine nichtlineare, narrative Komponente, die dem Benutzer wahlweise einen Einfluss auf die erzählte Geschichte mittels Branching oder einer morphologische Funktionsauswahl gewährt. Die Modellierung der Geschichte erfolgt mittels eines expliziten Modells, welches durch eine Story Engine ausgeführt werden kann.
- Eine konversationale Interaktionskomponente, die eine Notation von konversationalen Aspekten - sowohl propositional als auch diskursbezogen - erlaubt. Die Konversationalen Aspekte orientieren sich dabei an der Funktion, die der System-repräsentierende Konversationsteilnehmer - der Actor - besitzt. Die Konversation wird durch die Inhalte der Kontext-gebenden Applikation angetrieben, findet also in Bezug zu den präsentierten Videos statt.
- Eine medienspezifische, direkt manipulative Interaktionsmethodik gewährleistet einen expliziten, temporalen, intramedialen Zugriff des Benutzers auf die Videopräsentation mittels Videoannotationen.
- Ein Applikationsspezifischer Content-Service ist durch einen erweiterten kommerziellen Videoserver gegeben.

Als Besonderheit des DIVA Systems ist hervorzuheben, dass es als Online-System zum Einsatz kommt und vollständig über einen Web-Browser lauffähig ist.

6.4 Zusammenfassung

Betrachtet man die drei Applikationen vor dem Hintergrund ihrer Anforderungen und Szenarien, so wird deutlich, dass die Anforderungen sehr unterschiedlich sind:

- Das Projekt EMBASSI ist im privaten, nichtarbeitsbezogenen Bereich angesiedelt und soll die Nutzung des heterogenen Geräteparks im Privathaushalt standardisieren. Eine Assistenzfunktion ist aufgrund der geringen eigenen Intelligenz der Geräte notwendig, während ein direkt manipulativer Zugriff auf die Geräte - soweit möglich - eher vermieden wird. Zum besseren Verständnis der Benutzer soll ein anthropomorpher Avatar die Assistenzfunktion konversational vermitteln.
- Das Projekt MAP ist im Arbeitsumfeld angesiedelt, es hat als Ziel eine Erweiterung des traditionellen Arbeitsplatzes über Assistenz und Delegationsfunktionen, die dem Benutzer mittels Agenten über eine anthropomorphe, konversationale Schnittstelle zur Verfügung gestellt werden. Die Agenten selbst realisieren einen direkt manipulativen Zugriff auf Medien, falls sie diese Zugriffsart wünschen.
- DIVA ist ein System, das im Entertainment-Bereich angesiedelt ist. DIVA ermöglicht die interaktive Narration von Video. Dem Publikum des Videos wird sowohl eine Konversationale als auch eine direkt manipulative Interaktionsmöglichkeit geboten.

Trotz dieser unterschiedlichen Anforderungen sind die Mittel, mit denen die Ziele der szenarischen Vorgaben angegangen werden, in Teilbereichen identisch.

Die Forderung nach einer anthropomorphen Benutzerschnittstelle, welche eine Assistenz- und Delegationsfunktion für den Benutzer realisiert, ist ähnlich realisiert. Durch den Versuch einer konversationalen Modellierung kommen alle Systeme zum Ziel, wobei die Art der konversationalen Modellierung stark unterschiedlich ist. So wird im Projekt EMBASSI eine zentrale Dialogmodellierung genutzt, welche den sprachbasierten Dialog abbildet. Zusätzliche konversationale Elemente werden erst nach der Modalisierung des Dialoges - zum Beispiel durch einen Avatar - hinzugefügt. Diese konversationalen Elemente sind zudem direkt auf Speech Acts bezogen und im eigentlichen Sinne keine Erweiterung des Systems, sondern eine Funktion mit dem Speech Act als Eingabe und einem Verhalten als Ausgabe. Dies unterscheidet sich vom MAP System - dort wird die inhaltsgebende Dialogstruktur tatsächlich um grundlegende konversationale Aspekte erweitert, durch einen Automatismus, der in einer Conversation Engine abgebildet wird. Der Effekt ist ein zusätzliches Informationsangebot durch konversationales Verhalten des Systems. Belegen EMBASSI und MAP in dieser Bewertung die Extrema - EMBASSI durch eine simple Abbildung von konversational interpretierten Speech Acts, MAP durch eine konversationale Erweiterung neben den Inhaltselementen - so kann das DIVA System in der Mitte zwischen den beiden Extremen angeordnet werden. Im DIVA System wird zwar eine konversationale Modellierung vorgenommen, jedoch mit den Inhaltsinformationen der Konversation gemischt notiert. Die konversationale Ausformulierung obliegt somit dem Autoren einer Konversation. Beurteilt nach den Vorgaben für konversationale Modellierung, wie sie in dieser Arbeit aufgestellt sind - dies ist die explizite Formulierung von konversationalen Aspekten sowie die Trennung von konversationaler Modellierung und Inhalt - kann jede Applikation zwar das Konversationale Benutzerinterface als solches erreichen, die explizite Formulierung von konversationalen Aspekten findet jedoch nur im Projekt MAP und DIVA statt, die Trennung von Konversation und Inhalt wird nur in MAP durchgängig verwirklicht.

Die Systeme realisieren jeweils eine medienspezifische, direkt manipulative Interaktionsmöglichkeit für den Benutzer. Jedoch gewährt EMBASSI den jeweiligen Assistenten keine applikations-spezifische Lösung, der Kontakt mit dem Benutzer führt immer über das EMBASSI Assistenz-System. MAP und DIVA hingegen lösen die direkt manipulative Interaktion durch spezifische Anwendungen der jeweiligen, kontext-gebenden Applikation. Da ausschließlich DIVA auf eine Präsentation von Video optimiert ist, werden nur in DIVA die Konzeptvorgaben für eine direkte Manipulation von Video eingehalten - dies ist die Möglichkeit des Benutzers, auf explizite, temporale und intramediale Annotation des Videos zu interagieren. Das MAP System kann jedoch entsprechende Applikationen in Form von Agenten einbinden und ist somit potenziell in der Lage, diese Anforderung zu erfüllen.

Die narrative Komponente des Konzepts - die Forderung nach einer nichtlinearen Narration von Geschichten mittels kontinuierlicher Medien - wird von dem DIVA System realisiert. Tatsächlich ist in DIVA eine nichtlineare Zusammenstellung von Story Elementen - Video-Clips - gegeben. Diese Zusammenstellung kann sowohl händisch (manuell) - durch einen Autoren - als auch mittels einer morphologischen Beschreibung der Szenen und deren Verar-

beitung anhand eines morphologischen Modells mittels einer Story Engine geschehen. Durch die Erweiterung des selektiven Video-Clip Ansatzes um eine generative Komponente, den animierten Avatar, erfüllt das DIVA System die Forderung nach einem kombiniert selektiven/generativen Narrationsansatz. Die Projekte MAP und EMBASSI gewähren ebenfalls eine videobasierte Narration von Geschichten, sie modellieren diese jedoch nicht als zentrales Element der Inhaltsdarbietung sondern als eine von vielen möglichen Komponenten. Dies liegt in den Anforderungen der jeweiligen Szenarien begründet, die durch MAP und EMBASSI abgedeckt werden - sie sind erheblich breiter gefasst als das DIVA Szenario und umfassen zusätzliche Dienste (EMBASSI z.B. die Bedienung einer Haussteuerung oder eines Autoradios, MAP z.B. die Nutzung von Terminvereinbarungsagenten), die keine narrative Komponente vorsehen. Im Einzelfall ist auch die konversationale Unterstützung von kontinuierlichen Präsentationen gegeben. EMBASSI bietet einen Assistenten zur Auswahl von Funktionen rund um Fernsehen an - z.B. Programmwahl, Programmaufnahme. MAP bietet einen Videokonferenzagenten an, der sowohl die Konferenz selbst als auch die Annotation der Konferenz durch den Benutzer erlaubt.

Die generischen Systeme EMBASSI, MAP und DIVA werden im folgenden Kapitel in Bezug auf Anwendungen und Resultate geprüft. Diese Verifikation wird zur Bewertung der in Kapitel 1.1 aufgestellten Arbeitshypothesen herangezogen.

7. ANWENDUNGEN UND RESULTATE

Die generischen Applikationen EMBASSI, MAP und DIVA sind in verschiedenen Szenarien eingesetzt und evaluiert worden. Die Evaluation der Applikationen erfolgt im Hinblick auf die Anwendungsvorgaben und Anwendungsziele - aus diesem Grund werden die einzelnen Anwendungen im Kontext Ihrer Anwendungsszenarien beschrieben und die entsprechende Evaluation vor dem Hintergrund der jeweiligen Anwendungsszenarien diskutiert. Im Anschluss erfolgt eine Diskussion der Resultate im Hinblick auf das in Kapitel 5 vorgestellte Konzept der interaktiven Narration von kontinuierlichen Medien. Die Evaluation der Anwendungen folgt verschiedenen Prinzipien. So wurde im Projekt EMBASSI ein Nutzertest nach der Fertigstellung des Systems durchgeführt, das MAP Projekt wurde durch ein Usability Labor psychologisch überprüft und die DIVA Studie anhand von verschiedenen Studien vorbereitend initiiert.

7.1 EMBASSI Anwendungen

Die Diskussion der Evaluation des EMBASSI Systems, vgl. [HK01], erfolgt in dem Bereich A-PHH (Anwendung Privathaushalt), da dieser Bereich die Elemente, die für eine anschließende Diskussion des Gesamtkonzepts erforderlich sind, beinhaltet. Die Anwendung wird darauf geprüft,

- ob der Einsatz eines virtuellen Helfers tatsächlich einen Vorteil bei der Bedienung des EMBASSI Systems verspricht.
- ob der Einsatz eines virtuellen Helfers dem Benutzer eine (subjektiv) natürlichere Umgangsform mit dem EMBASSI System erlaubt.
- Ob der Einsatz eines virtuellen Helfers den Benutzer zur konversationalen Interaktion mit dem EMBASSI System stimuliert.

Die folgenden Unterkapitel beschreiben die entsprechenden Szenarien und Tests des EMBASSI Systems. Anschließend werden die Tests kritisch diskutiert und bewertet.

7.1.1 Szenario

Folgende Anwendungsszenarien sind im Bereich A-PHH evaluiert:

- die Bedienung des Fernsehers
- die Bedienung eines Videorecorders
- die Helligkeitsanpassung in der Wohnung
- das Besorgen und das Anschauen eines Videos aus dem Internet

Im den folgenden Absätzen sind die Szenarien weiter beschrieben.

Das Szenario *Bedienung eines Fernsehers* nutzt wichtige Aspekte der multimodalen Interaktion mit dem Benutzer. So definiert der Benutzer seine Aufforderung an das System mit Hilfe von Gestik und Sprache, anstatt seine Befehle direkt manipulativ über Schalter und andere Eingabemöglichkeiten des Fernsehers zu kommunizieren. Das EMBASSI System hat die Aufgabe, die Eingaben des Benutzers im Sinne der Ziele des Benutzers zu interpretieren und anschließend die Benutzerziele tatsächlich durchzuführen.

Das Szenario *Bedienung eines Videorekorders* stellt eine weitere, scheinbar einfache Delegation an das EMBASSI System dar. Eine sprachliche Eingabe des Benutzers -etwa *Ich möchte heute Abend den Krimi auf ARD aufnehmen* stellt das System vor die Aufgabe, ein komplexes Ziel des Benutzers, ohne weitere Hinweise formuliert, zu interpretieren und in die verschiedenen durch das System durchführbaren Aufgaben zu zerlegen - dies nur unterstützt durch unbedingt notwendige Nachfragen unter Vermeidung von unnötigen Ausgaben. Offensichtlich lassen sich für EMBASSI die verschiedenen Grundinformationen nach Zeit - *heute Abend* - nach Programmwahl - *ARD* - und dem Sendungstypen - *Krimi* - mit der Aktion *aufnehmen* assoziieren. Um diese allerdings in durchführbare Aufgaben umzusetzen, ist ein Wissen über das Gesamtsystem und den Benutzer notwendig. Fragen nach der Einsatzbereitschaft der verwalteten Geräte sind genauso zu klären wie die Überprüfung der Sinnhaftigkeit der Benutzeraussage (gibt es einen Krimi um diese Uhrzeit?). Im gegebenen Falle werden mögliche alternative Vorschläge des Systems generiert.

Das Szenario *Helligkeitsanpassung* demonstriert die Haussteuerungsfunktionalitäten des EMBASSI Systems. Das Szenario beschreibt einen Benutzer, der während des Fernsehens durch die einfallende Sonne gestört wird - er bittet das EMBASSI System zur Regulierung der Helligkeit im Zimmer durch die simple Aufforderung *heller*. Mit Hilfe des im System gespeicherten Wissens über verfügbare Geräte und deren Einstellungsmöglichkeiten sowie über Präferenzen des Benutzers wird vom EMBASSI System eine geeignete Strategie ausgearbeitet und angewandt, um die Befehle des Benutzers umzusetzen. Dies kann in diesem Szenario zum Beispiel durch ein Anheben der Helligkeit des Fernsehbildes oder durch ein Abdunkeln des Raumes durch Herablassen der Rollos erfolgen.

Das Szenario *Besorgen und das Anschauen eines Videos aus dem Internet* demonstriert den Einsatz eines elektornischen Einkauf-Assistenten, der auf einen Filmwunsch des Benutzers reagiert. Der Filmwunsch wird in eine System-verständliche Form umgesetzt und im Anschluss der Einkaufs-Assistent mit dem Einkauf des Films beauftragt (zu Bestpreiskonditionen!). Nach erfolgreicher Ausführung des Einkaufs wird der Film für den Benutzer auf dem Bildschirm des Fernsehers präsentiert.

Das Szenario *Besorgen und das Anschauen eines Videos aus dem Internet* wird zum besseren Verständnis bezüglich des Dialogmanagements vertieft. Als Hauptziel der Diskussion wird die semantische Umsetzung von delegativen Benutzereingaben in entsprechende Systemassistenten erläutert. Die technische Umsetzung des Szenarios wird bezüglich der Geräte-Ebene bzw. der Geräte-Ansteuerung nicht diskutiert.

Die semantische Umsetzung erfolgt durch Zuordnung der vom System erkannten Benutzerziele zu Assistenten, die diese Ziele erreichen. Dazu muss in einem ersten Schritt der geeignete Assistent durch das Dialogsystem identifiziert werden. Das Szenario *Besorgen und das Anschauen eines Videos aus dem Internet* nutzt den Shopping-Assistenten, da das Ziel des Benutzers der Kauf eines bestimmten Films (Videos) und das anschließende Ansehen des Filmes durch den Benutzer ist. Dieses Ziel wird - als KQML Nachricht formuliert - an den Assistenten delegiert. Der entsprechende Assistent - in diesem Fall ein mobiler Agent auf der Agentenplattform SeMoA (Secure Mobile Agents) - informiert sich selbstständig über die Adressen verschiedener Online-Anbieter, migriert zu den Anbietern vor Ort, verhandelt und kauft schließlich den Film. Der Assistent veranlasst die Sicherung des Filmes durch den Massenspeicher-Assistenten im EMBASSI System des Nutzers. Anschließend wendet er sich mit einer Erfolgsmeldung an den Dialogmanager. Der Dialogmanager erkennt, dass das erste Teilziel - der Kauf des Films - erreicht ist und veranlaßt das zweite Teilziel - die Vorführung des Filmes - durch den MPEG-Assistenten. Der MPEG-Assistent formuliert entsprechende Anfragen an Massenspeicher-Assistenten und sonstige Geräte-Assistenten, so dass der Film schließlich abgespielt wird.

7.1.2 Anwendungstest

Die Anwendungstests¹ sind in direktem Zusammenhang mit den grundsätzlichen Anforderungen an das EMBASSI System zu sehen. Ziel des Projektes ist es, die Mensch-Technik Interaktion positiv zu beeinflussen, daher die technischen Geräte im Alltag des Benutzers für diesen *effizient*² nutzbar zu machen - dies auf dem Schwerpunkt Unterhaltungselektronik.

Darüberhinaus war das Ziel der Untersuchung die Auswirkung des virtuellen Helfers (Avatars) des EMBASSI Systems auf natürlichsprachliche Interaktion bzw. die Auswahl von Eingabemodalitäten. Als Testszenario ist das in Kapitel 7.1.1 angegebene Szenario *Bedienung eines Videorekorders* gewählt.

¹ Die psychologische Untersuchung des Projektes wurde durchgeführt von der Universität Köln, Psychologisches Institut, sowie der Universität Berlin, Zentrum Mensch Maschine Systeme. Die entsprechenden Ergebnisse wurden veröffentlicht von Dr. Nicole Krämer und Julia Nitschke, vgl. [KN01]. Die Untersuchungen diskutieren die adäquate Unterstützung von Benutzern, je nach Benutzereigenschaften, nach Aufgaben der Benutzer und der Situation der Nutzung. Konkret wurde die Akzeptanz und Effizienz der angebotenen Ausgabemodalitäten des EMBASSI Systems untersucht.

² Der Begriff Effizient wird dabei im folgenden numerisch definiert.

7.1.2.1 Methodik und Durchführung

Zur Analyse des EMBASSI Systems wurde ein statistisches Testverfahren (Varianzanalyse³) benutzt.

Die Methodik des EMBASSI Tests wird im folgenden Beschrieben:

- 65 Personen führten prototypische Bedienaufgaben an einem kombinierten EMBASSI TV-Video-System durch.
- Die Bedienung erfolgte per Fernbedienung, GUI oder Spracheingabe.
- Aufgabe: Eine Fernsehsendung nach Wahl bzw. eine vorgegebene Sendung sollte ausgewählt und aufgezeichnet werden.

Aufgabe 1 Die erste Aufgabe ist die Aufzeichnung einer Sendung nach Wahl.

Aufgabe 2 Die zweite Aufgabe ist die Aufzeichnung einer Sendung nach Wahl.

Aufgabe 3 Die dritte Aufgabe ist die Aufzeichnung der Sendung *Boulevard Bio*.

- Randbedingungen:
 - Die Programmierung der Sendung erfolgt automatisch nach Auswahl der Sendung, wenn der Befehl zur Aufzeichnung gegeben wird.
 - GUI-Listen konnten sortiert werden nach Uhrzeit, Sendern und Genres.
 - Die Anzahl der auswählbaren Sendungen entspricht einem durchschnittlichen Abendprogramm.
 - Sowohl Spracheingabe als auch GUI durften jeweils benutzt werden.

Wie jeder psychologische Test besitzt auch der hier beschriebene eine Anzahl abhängiger und unabhängiger Variablen, welche den Test beeinflussen. Die unabhängigen Variablen des Test sind in der Variation der Ausgabemodalitäten zu sehen:

Versuchsgruppe 1 Diese Gruppe bearbeitete die ersten beiden Aufgaben mittels der GUI. Die dritte Aufgabe durfte wie durch Versuchsgruppe 4 bearbeitet werden.

Versuchsgruppe 2 Diese Gruppe bearbeitet die ersten beiden Aufgaben mit Hilfe der GUI und mit Hilfe von Sprachausgabe. Die Sprachausgabe informierte über Systemzustände. Die dritte Aufgabe durfte wie durch Versuchsgruppe 4 bearbeitet werden.

³ Mit einer Varianzanalyse wird statistisch überprüft, ob Unterschiede zwischen Gruppen zufällig aufgetreten sind, oder ob wirkliche Unterschiede zwischen diesen Gruppen existieren. Dies geschieht, indem man die Variation zwischen den Gruppen mit der Variation innerhalb der Gruppen vergleicht. Zu diesem Zweck werden abhängige und unabhängige Variablen notiert und die Fragestellung geprüft, wie sich eine oder mehrere unabhängige Variablen auf eine oder mehrere abhängige Variablen auswirken. Die statistischen Grundlagen zur Varianzanalyse sind u.a. durch Henze, vgl. [Hen97], und durch Fahrmeir, vgl. [FKPT98], beschrieben.



Fig. 7.1: EMBASSI Experiment - links GUI (Grundig), rechts virtueller Charakter (ZGDV), [KN01]

Versuchsgruppe 3 Diese Gruppe bearbeitete die ersten beiden Aufgaben mit Hilfe der GUI und Sprachausgabe, zusätzlich wurde ein virtueller Charakter (der EMBASSI Avatar) auf der rechten Seite des Bildschirms eingeblendet, welcher die Sprachausgabe lippen-synchron begleitete, siehe Abbildung 7.1. Die dritte Aufgabe durfte entsprechend Versuchsgruppe 4 bearbeitet werden.

Versuchsgruppe 4 Diese Gruppe durfte für jede Aufgabe eine der drei oben beschriebenen Varianten wählen.

Die Varianzanalyse prüft, ob die verschiedenen Gruppen sich unterschiedlich verhalten. Dabei wird angenommen, dass die Gruppen sich nicht abweichend Verhalten dürften, daher jede Gruppe im Mittel das selbe Verhalten zeigt.

In der Varianzanalyse geht es allgemein um die Fragestellung, wie sich eine oder mehrere unabhängige Variablen auf eine oder mehrere abhängige Variablen auswirken. Die abhängigen Variablen des Tests dienen der Erfassung der Benutzereffizienz. Die Benutzer selbst sind im Vorfeld im Hinblick auf personale Angaben (Alter, Geschlecht, Technik-Erfahrung) überprüft worden. Die abhängigen Variablen sind im folgenden notiert.

- Bearbeitungsschritte und Schnelligkeit (Kontrolliert durch LogFile-Daten): Nachvollziehen der Vorgehensweise der Benutzer (z.B. Zeiten und Bedienschritte).
- Verhalten der Benutzer (parallele Video und Ton-Aufzeichnung)
- Zufriedenheitsmaß (anschließender Fragebogen): Erfassung der subjektiven Akzeptanz der Benutzer im Hinblick auf Ausgabemodalitäten. Hierzu wurden Fragen nach subjektivem Empfinden und Bewertung der Effizienz des Gesamtsystems gestellt.

Der Fragebogen teilte sich in einen Bereich während des Test und eine post-hoc Bewertung, die jeweils verschiedene Faktoren enthält. Die Varianzaufklärung der Faktoren gibt an, wie die Abweichung der angenommenen Testmittelwerte prozentual vom entsprechenden Faktor abhängt.

Der parallel zum Test auszufüllende Fragebogen enthält drei Faktoren, die sich aus Untermengen der in Abbildung 7.2 gezeigten Verben der Y-Achse zusammen setzen:

- Interesse/Aufmerksamkeit (Varianzaufklärung: 21,75%)
- Verärgerung (Varianzaufklärung: 17,36%)
- Entspannung/positive Gefühle (Varianzaufklärung: 11,81%)

Der post-hoc Fragebogen enthält ebenfalls drei Faktoren:

- Evaluation/empfundene Effizienz (Varianzaufklärung: 32,89%)
- Spaß (Varianzaufklärung: 20,08%)
- Kontrolle/Schwierigkeit (Varianzaufklärung: 19,24%)

Tab. 7.1: EMBASSI Versuchsplan mit abhängigen und unabhängigen Variablen, [KN01]

Anzahl der Probanden	18	16	15	16
Geschlecht	9 männlich / 9 weiblich	7 männlich / 9 weiblich	7 männlich / 8 weiblich	6 männlich / 10 weiblich
Unabhängige Variable	GUI	Sprachausgabe	Virtueller Helfer	freie Wahl
Abhängige Variablen	Empfinden während der Interaktion Bewertung Empfundene Effizienz Anzahl der Spracheingaben Zeit bis zur Lösung Anzahl der Schritte Getroffene Wahl hinsichtlich Ausgabemodalitäten			
Moderatorvariablen	Alter Geschlecht Expertise			

Die abhängigen und unabhängigen Variablen werden in Tabelle 7.1 veranschaulicht. Die Durchführung der Tests fand wie folgt statt:

1. Die Benutzer wurden mit der Bedienung des Systems vertraut gemacht, dies erfolgte Anhand von Beispielen und Übungen.

2. Die ersten beiden Aufgaben wurden bearbeitet.
3. Die Benutzer der Gruppen 1, 2 und 3 bearbeiteten den Fragebogen bezüglich der Akzeptanz des Systems.
4. Die Benutzer der Gruppen 1, 2 und 3 durften für die dritte Aufgabe eine Modalität wählen.
5. Benutzer der Gruppen 1 bis 4 füllten den Fragebogen aus, falls die Modalität gewechselt wurde.

Die Ergebnisse des Tests sind in folgenden Kapitel beschrieben.

7.1.2.2 Ergebnisse

Die Ergebnisse des Tests teilen sich in Werte bezüglich der Heterogenität der Stichprobe und der Moderatorvariablen (Zusammensetzung der Benutzergruppen) sowie das Akzeptanzverhalten und die Effizienz der Benutzer bezüglich des EMBASSI Systems.

Es kann bezüglich der Stichprobe eine heterogene Struktur von 29 Männern und 36 Frauen belegt werden, die im Durchschnitt 39,28 Jahre alt waren und in eine Expertengruppe (33) und eine Novizengruppe (32) eingeteilt werden konnten. Das Geschlecht (als Moderatorvariable) zeigte einen sehr geringen Einfluss auf das Testergebnis - insbesondere die Bewertung des Systems fiel bei Frauen insgesamt besser aus als bei Männern. Die sonstigen Moderatorvariablen ließen keine Schlüsse auf eine Abhängigkeit bezüglich Performance der Benutzer oder Neigung zu bestimmten Modalitäten zu.

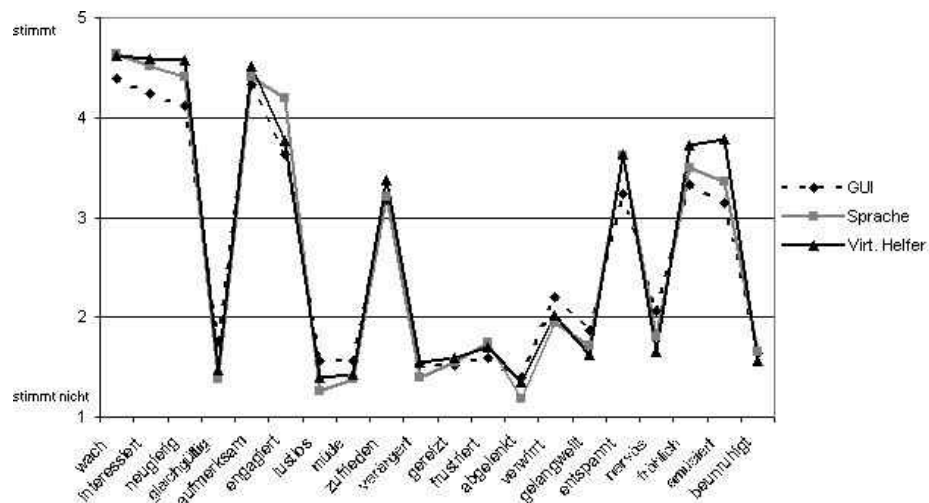


Fig. 7.2: EMBASSI Experiment - Mittelwert zum Faktor *Interesse/Aufmerksamkeit*, [KN01]

Die Akzeptanz des Systems wurde für das Wahlverhalten der Benutzer objektiv, für die Meinung der Benutzer subjektiv erfasst. Dabei waren folgende Aspekte feststellbar:

- **Objektiv:** Der virtuelle Helfer wurde von den Benutzern jeder Versuchsgruppe überdurchschnittlich oft als Ausgabemodalität gewählt. Auch nach zwischenzeitlichem Wechsel der Ausgabemodalität wurde der virtuelle Helfer oft wieder benutzt.
- **Subjektiv:** Hinsichtlich der Bewertung der Modalitäten lassen sich tatsächlich keine signifikanten Unterschiede feststellen. Einzig für den Faktor *Interesse/Aufmerksamkeit* wird eine Signifikanz festgestellt, welche auf eine niedrigere Aufmerksamkeit bei der reinen GUI Anwendung zurückzuführen ist, siehe Abbildung 7.2.

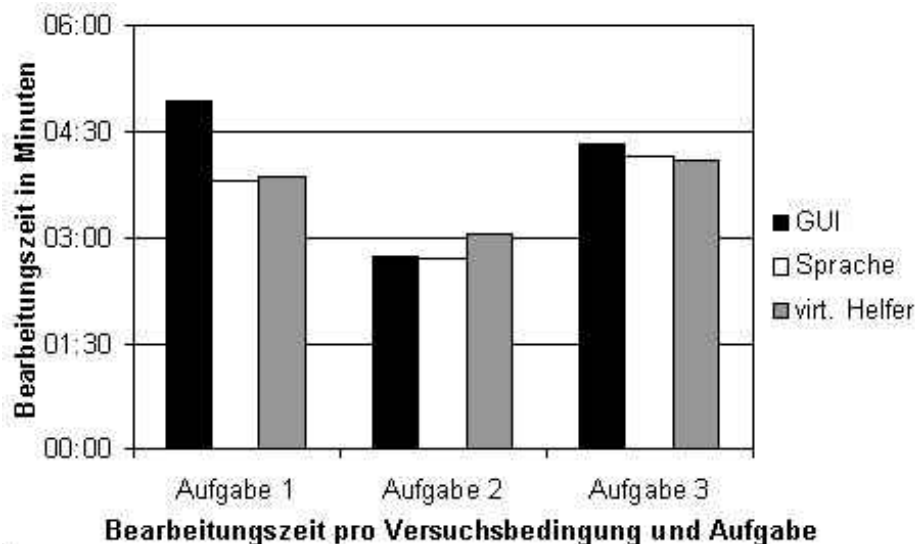


Fig. 7.3: EMBASSI Experiment - Durchschnittliche Bearbeitungszeit, [KN01]

Die Effizienz des EMBASSI Ansatzes wurde anhand der Logfile Daten überprüft mit dem Ergebnis, dass die Bearbeitungszeit pro Aufgabe nicht in Abhängigkeit zur gewählten Ausgabemodalität steht. Lediglich die rein GUI basierten Möglichkeiten der Versuchsgruppe 1 führten bei Aufgabe 1 zu einer signifikanten Erhöhung der Anzahl der Bearbeitungsschritte. Dies relativierte sich jedoch wieder bei Aufgabe 2, siehe die Abbildungen 7.3 und 7.4.

Das Eingabeverhalten der Benutzer zeigte die deutlichsten Unterschiede. Hier lässt sich feststellen, dass in diesem Test das Verhältnis von multimodaler, anthropomorpher Ausgabe zu natursprachlichem Eingabeverhalten proportional ist, siehe Abbildung 7.5. Die Testergebnisse zeigen einen Unterschied in der Signifikanz: Signifikant ($se=0,95$; $p=0,014$ - se entspricht der Signifikanz, p der Wahrscheinlichkeit einer Abweichung des Signifikanz-Wertes) mehr natursprachliche Äußerungen bei virtuellem Helfer im Vergleich zu GUI - dies bei Analyse von einzelnen Aufgaben, wie auch bei der übergreifenden Analyse aller Test-Durchgänge.

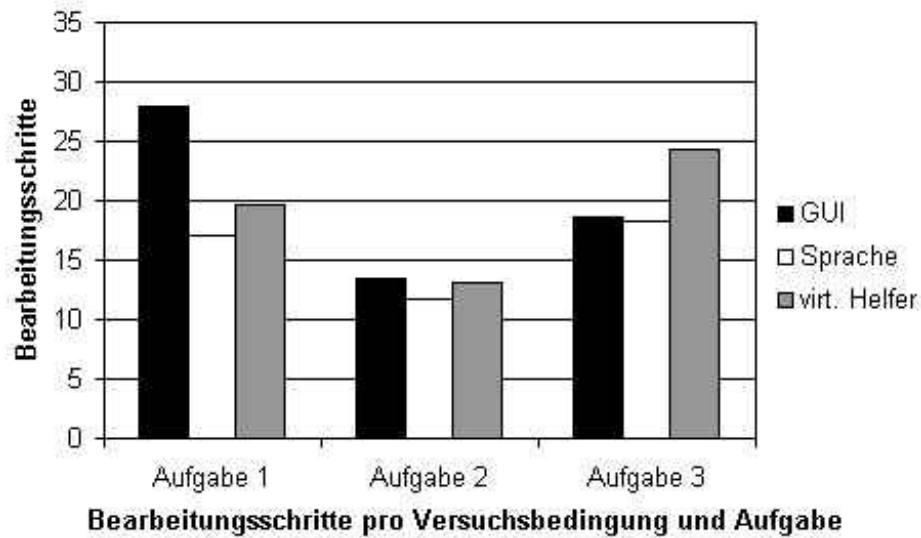


Fig. 7.4: EMBASSI Experiment - Durchschnittliche Anzahl Bearbeitungsschritte, [KN01]

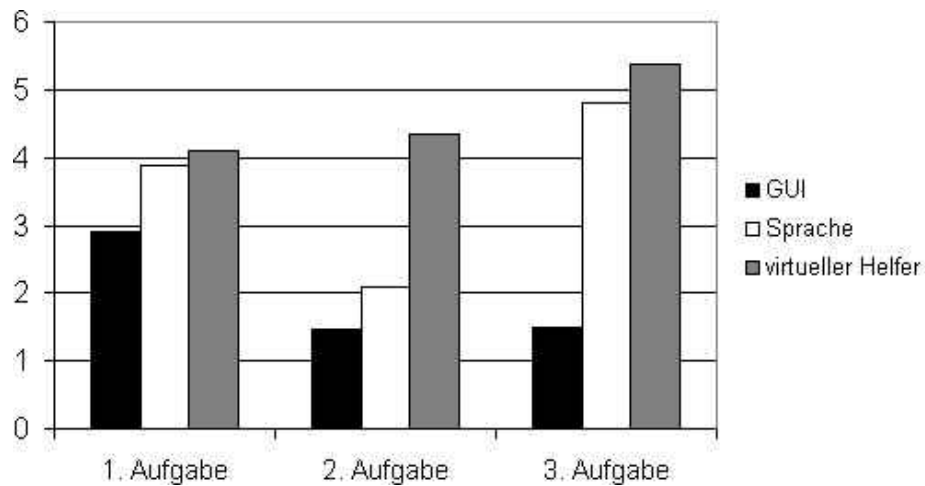


Fig. 7.5: EMBASSI Experiment - Anzahl natursprachliche Eingaben, [KN01]

7.1.3 Diskussion

Die Ergebnisse des Test werden in diesem Kapitel hinsichtlich der Fragestellungen interpretiert, wie sie für das EMBASSI System gestellt sind.

Vorteil Bedienung Der Einsatz eines virtuellen Helfers bietet nur einen hohen subjektiven Vorteil bei der Bedienung des EMBASSI Systems. Tatsächlich sind die Benutzer des Systems - subjektiv - interessierter und aufmerksamer, wenn der virtuelle Helfer zum Einsatz kommt.

Natürliche Umgangsform Der Einsatz eines virtuellen Helfers erlaubt dem Benutzer eine (subjektiv) natürlichere Umgangsform mit dem EMBASSI System, was durch die positive Bewertung des virtuellen Helfers durch die Benutzer verifiziert wird.

Konversationale Interaktion Der Einsatz eines virtuellen Helfers zur Interaktion mit dem EMBASSI System stimuliert tatsächlich den Einsatz von Konversation zur Interaktion des Benutzers mit dem System. Dies ist am deutlichsten an dem Resultat zu erkennen, welches die Abbildung eines virtuellen Helfers und die damit verbundene Erhöhung der natürlichsprachlichen Eingabe postuliert.

Krämer und Nitschke, vgl. [KN01], kritisieren die Aufgabenstellung des Tests im Hinblick auf mangelnde Komplexität der Aufgabe - sie folgern aus der erhöhten Anzahl der Bedienschritte von Versuchsgruppe 1 bei der ersten Aufgabe, dass eine komplexere Aufgabe die Performance-Unterschiede der Benutzer in Relation zu den gewählten Modalitäten deutlich in Richtung des virtuellen Helfers verschoben hätten.

Insgesamt zeigt der Test, dass das Ziel der - subjektiv wie objektiv - besseren Bedienbarkeit durch Assistenz und Delegation des EMBASSI Systems erreicht wurde. Die Ergebnisse des Tests sind durchweg auf die konversationalen Anteile des Grundkonzepts (Kapitel 5) dieser Arbeit übertragbar, da diese durch das EMBASSI System - zumindest im Testszenario - entsprechend modelliert sind.

7.2 MAP Anwendungen

Die Diskussion des MAP Systems erfolgt anhand einer psychologischen Beurteilung der Akzeptanz und Akzeptabilität des MAP Systems⁴ im Hinblick auf Aufgabendelegation und Assistenz durch den Benutzer mittels des MAP User Interface Agenten. Dieses Vorgehen erfolgt im Wissen, dass eine wichtige Voraussetzung für die Akzeptanz eines jeden Produkts die Möglichkeit des Benutzers ist, dieses Produkt tatsächlich zu gebrauchen. Die Gebrauchstauglichkeit eines Produkts impliziert dabei sowohl die *Nützlichkeit* als auch die *Bedienbarkeit* des Produkts. Um einen nachvollziehbaren Wert für die Gebrauchstauglichkeit eines Produkts zu erhalten, werden folgende Punkte untersucht:

Effektivität Bis zu welchem Grad werden die Arbeitsziele erreicht?

Effizienz In welcher Relation stehen Arbeitsaufwand und erreichte Ziele?

Akzeptanz Wie zufrieden ist der Benutzer des Produktes?

Als prototypisches Szenario ist das Bauwesen ausgewählt [Han00]. Arbeiten wie Projektsteuerung, Bauüberwachung bzw. in Architekturbüros die komplette Überwachung der Bauphasen Planung, Bau und Nutzung sind mit einem erheblichen Wechsel zwischen Phasen der Bürotätigkeit und Tätigkeiten vor Ort (auf der Baustelle) und dadurch mit einem hohen

⁴ Diese Beurteilung wurde durch das Siemens Usability Labor, vgl. [SH01], durchgeführt

Koordinationsaufwand verbunden. Die Hilfestellung des MAP mittels des User Interface Agenten kann durch den hohen Anteil an delegativen Aufgaben effizient überprüft werden. Im Folgenden werden die Grundlagen der Usability-Beurteilung notiert sowie die Usability Beurteilung selbst vorgestellt. Diese Beurteilung wird im Anschluss diskutiert.

7.2.1 Voraussetzung

Die klassische Aufgabe eines intelligenten Software Agenten liegt in der Reduzierung von komplexen Aufgaben auf eine simple Delegation der Aufgabe durch den Benutzer, der sich um die Details der Aufgabe keine weiteren Gedanken machen möchte. Aus diesem Grund muss der intelligente Agent eine Prognose über das Handlungsziel des Benutzers anstellen, um ihm die erwartete Hilfestellung tatsächlich gewähren zu können. Aus Sicht der Usability-Bewertung ist eine Beurteilung der durch den Benutzer wahrnehmbaren Aspekte - Intelligenz, Adaptivität, Proaktivität und Lernfähigkeit - von intelligenten Agenten notwendig, dies in Relation zu klassischen Usability Prinzipien, vgl. [Hö98], wie Kontrolle, Vorhersagbarkeit und Transparenz, siehe Tabelle 7.2.

Tab. 7.2: Usability Prinzipien nach Höök [Hö98]

Prinzip	Bedeutung
Kontrolle	Der Benutzer hat jederzeit die Möglichkeit, eine Aktion des Systems zu überstimmen/zu kontrollieren
Vorhersagbarkeit	Der Benutzer weiß, nach welchen Regeln/Prinzipien/Algorithmen eine Entscheidung des Systems getroffen wird
Transparenz	Der Benutzer muss Kenntnis darüber haben, welche Informationen/Eingaben zu welchem Ergebnis des Systems führen

Weitere Prinzipien sind in ISO 9241-10, vgl. [Sch01], [ISO98], als die sieben Grundsätze für die Gestaltung und Bewertung eines Dialoges definiert worden:

- Aufgabenangemessenheit
- Selbstbeschreibungsfähigkeit
- Steuerbarkeit
- Erwartungskonformität
- Fehlerrobustheit
- Individualisierbarkeit
- Lernförderlichkeit

Diese Grundsätze sollten als allgemeine Leitlinien bei der Gestaltung und Bewertung von Dialogsystemen angewandt werden. Einschränkend wirkt dabei, dass die Umsetzung der einzelnen Grundsätze der Dialoggestaltung von den Merkmalen des Benutzers, für den das Dialogsystem gedacht ist, den Arbeitsaufgaben und der Arbeitsumgebung abhängt. Ganz im Sinne einer intuitiven Konversation, wie sie in MAP geplant ist, verfolgen die Dialoggrundsätze einen benutzerorientierten Ansatz an die Entwicklung eines konversationalen Systems. Das Ziel ist die Vereinfachung und Verbesserung der Benutzung von Dialogsystemen durch die gestiegene Qualität des Dialogs. Die Dialoggrundsätze werden in einfachen Begrifflichkeiten definiert, um den Dialog zwischen Konversationssystem und Benutzer zu beschreiben. Besondere Berücksichtigung müssen jedoch auch die psychologischen Merkmale des Benutzers finden, welche beispielhaft im folgenden aufgelistet sind:

- Aufmerksamkeitsspanne
- Grenzen des Kurzzeitgedächtnisses
- Lerngewohnheiten
- Grad an Erfahrung
- das innere Modell des Benutzers von der zugrundeliegenden Struktur und dem Zweck des Dialogsystems, mit dem der Benutzer arbeiten wird.

Bei Dialogsystemen kommt es häufig vor, dass Benutzer mit unterschiedlichen Fähigkeiten zu unterstützen sind. Diese Fähigkeiten hängen ab von Faktoren wie:

- Vorkenntnis über den Arbeitsablauf.
- Vertrautheit mit dem Dialogsystem.
- Häufigkeit der Benutzung des Dialogsystems oder der Anwendung.

Bei Mensch-Maschine-Interaktionen in MAP tritt der Benutzer in einen Dialog mit dem Konversationssystem des User Interface Agenten, um einen bestimmten Arbeitsablauf zu erledigen. An dieser Stelle ist die Technik der Dialoggestaltung (zum Beispiel durch Diskursmodelle) von besonderer Bedeutung. Es kann an dieser Stelle notwendig sein, zuzüglich zur Konversation noch Direkt Manipulative Techniken zuzulassen. Die Kombination dieser Techniken gestaltet den Dialog, der den Benutzerbelangen, den Arbeitsaufgaben und der Arbeitsumgebung gerecht wird.

Für die kombinierte Interaktion des Benutzers mit dem User Interface Agenten und damit dem MAP System sind entsprechende Anforderungen und Überlegungen zur Gestaltung des Dialogs zu berücksichtigen. In ISO 9241-10 wird unter anderem hervorgehoben, dass die Grundsätze der Dialoggestaltung nicht unabhängig voneinander zu sehen sind. Es kann daher notwendig sein, einen Grundsatz höher zu bewerten als einen Anderen. Die Anwendbarkeit und die jeweilige Wichtigkeit hängen vom speziellen Anwendungsfall, von den Benutzergruppen und von der gewählten Dialogtechnik ab. Dies legt nahe, die folgenden Gesichtspunkte zu berücksichtigen:

- Ziele des Anwenders.
- Benutzerbelange der vorgesehenen Benutzergruppe.
- Aufgaben, die durch die Anwendung unterstützt werden sollen.
- Verfügbare Techniken und Mittel.

Es kann notwendig sein, bei der Anwendung der Grundsätze je nach Fall Prioritäten festzulegen.

Neben den (umfangreichen) Prinzipien der ISO 9241-10 kann die Usability Heuristik von Nielsen, vgl. [NM96], sehr effizient zur Bewertung von agentenbasierten Applikationen, vgl. Kaasinen [Kaa98], herangezogen werden, wobei die ursprüngliche Heuristik für die Gestaltung von direkt manipulativen User Interfaces aufgestellt wurde. Nielsen bediente sich einer Faktoranalyse von 249 Usability Problemen, um daraus 10 Heuristiken zu entwickeln. Die folgende Aufzählung notiert die Usability Heuristic von Nielsen sowie die Adaption durch Kaasinen.

1. Visibility of system status

Das System sollte dem Benutzer die vom System gerade vorgenommenen Arbeiten mit einem angemessenen Feedback zu einer angemessenen Zeit mitteilen.

Der Agent sollte den Benutzer darüber informieren, was er gerade an Arbeiten erledigt - dieses in angemessener Form zur angemessenen Zeit.

2. Match between system and the real world

Das System sollte für Ein- und Ausgaben Konzepte und Metaphern verwenden, welche dem Benutzer bekannt sind.

Der Agent sollte über eine personalisierte Schnittstelle verfügen, um dem Benutzer eine möglichst natürliche Ein- und Ausgabeschnittstelle zu bieten.

3. User control and freedom

Der Benutzer sollte die Möglichkeit haben Systemvorgänge vorzeitig abubrechen.

Der Benutzer muss die Kontrolle über die Arbeit des Agenten behalten - weiterhin sollte er die Möglichkeit besitzen, die Arbeit des Agenten zu unterbrechen und zu stoppen.

4. Consistency and standards

Ein- und Ausgaben sollten sich in jeder Situation an Konventionen halten, welche dem Benutzer bekannt sind.

Adaption sollte nicht die Konsistenz des Agentensystems untergraben. Auch personalisierte Ausgaben müssen eine Konsistenz einhalten.

5. Error prevention

Das System sollte Fehler vermeiden.

Speziell lernende Agenten sind nicht davor geschützt Fehlentscheidungen zu treffen. Der Benutzer muss deshalb jederzeit die Lösungsfindung und Lösungswege nachvollziehen können.

6. Recognition rather than recall

Objekte, Aktionen und Optionen müssen für den Benutzer explizit sichtbar sein. Eine längere Erinnerungszeit ist dem Benutzer i.A. nicht zumutbar.

Innovative Visualisierungsmethoden sollten dem Benutzer die Aktionen und Optionen der Agenten jederzeit visualisieren.

7. Flexibility and efficiency of use

Direkte Zugriffe auf Befehle können die Arbeit des Experten mit dem System extrem beschleunigen. Oft ausgeführte Aktionen sollten zu einem komplexen Befehl zusammengesetzt werden können.

Personalisierung kann das System effizient an den Benutzer und seine Bedürfnisse anpassen.

8. Aesthetic and minimalist design

Dialoge sollten mit einem Fokus auf wichtige Informationen geführt werden.

Personifizierte Agenten sollten die Kommunikation mit dem Benutzer auf wichtige Informationen fokussieren und dem Benutzer keine redundante oder (aus Benutzersicht) unnötige Kommunikation mit einem Pool von Agenten zumuten.

9. Help users recognize, diagnose, and recover from errors

Fehlermeldungen sollten dem Benutzer eine Möglichkeit geben, das Problem präzise zu erfassen und außerdem eine Lösung des Problems anbieten.

Durch Delegation und Assistenz in Agentensystemen können sehr komplizierte Fehler auftreten. Das Bearbeiten von Fehlern durch den Benutzer muss explizit im Agentensystem mit modelliert werden.

10. Help and documentation

Hilfesysteme sollten einfach zu durchsuchen und fokussiert auf Benutzerprobleme- und Aufgaben sein. Sie sollten konkrete Lösungsanweisungen mitführen und nicht zu kompliziert sein.

Der Benutzer sollte sich über die Möglichkeiten des Agenten bzw. des Agentensystems im klaren sein. Es ist ratsam, einen speziellen Hilfeagenten zu entwerfen.

Neben den allgemeinen Usability Anforderungen an die Schnittstelle zwischen Mensch und Computer werden spezielle, Agenten-zentrierte Sichtweisen aus Software-ergonomischer

Sicht spezifiziert. Dies bezieht sich in erster Linie auf die Aufgabenverteilung zwischen Mensch und Agent. Um ein Ziel zu erreichen, muss bei Mensch und Agent das selbe Verständnis dieses Ziels vorliegen - auch muss die Verteilung der Arbeitsschritte zwischen den Partnern zur Erreichung des Zieles definiert und durchführbar sein. Laurel [Lau90] listet eine Anzahl von Arbeiten, welche von Agenten für den Benutzer übernommen werden können - sei es durch fehlende Expertise oder Ressourcen beim Benutzer oder einfach durch Zeitmangel:

Information Navigieren und Browsen; Informationssuche; Sortieren und Organisieren; Filtern.

Arbeit Erinnern; Programmieren; Terminplanung; Anzeigen.

Lernen Training; Betreuen; Hilfesysteme.

Entertainment Gegner; Mitspieler; Darsteller.

Die Verteilung der Rollen zwischen Benutzer und Agent sind dabei fest definiert - der Benutzer bleibt immer der Teamleiter mit Verantwortungs- und Entscheidungskontrolle.

Bezüglich der Visualisierung - speziell im Kontext des MAP User Interface Agenten - impliziert der Begriff Agent eine für den Benutzer wahrnehmbare und damit sichtbare Gestaltung an der Bedienoberfläche⁵. Die häufig gewählte anthropomorphe Darstellung muss aufgrund Ihrer personifizierenden Darstellung aus ergonomischer Sicht hinterfragt werden - es könnte schließlich sinnvoller sein ausschließlich die zugrundeliegenden Techniken für den Benutzer sichtbar zu machen. So kann es bei besonders realistischer, menschähnlicher Darstellung eines Agenten zu Fehlschlüssen über die Fähigkeiten des Agenten kommen. Ben Shneiderman, vgl. [Shn97], etwa führt folgende Punkte als Gegenargumente an:

- anthropomorphe Repräsentanten erhöhen die Angst vor Computern.
- Agenten stehen im Konflikt zum Prinzip der Vorhersagbarkeit.
- Agenten verringern die Kontrolle des Benutzers über das System.
- Agenten untergraben die Verantwortlichkeit des Benutzers.

Wird der Agent trotz dieser Problematiken visualisiert, so sind die Designentscheidungen bezüglich des anthropomorphen Erscheinungsbilds und genereller Charakteristik des Agenten, vgl. [Kaa98], in Bezug auf aktiver (dominant, expertenhaft) oder passiver Agent (Untertan), Menschähnlich oder Comic-hafte Erscheinung, Anpassung an Kultur bez. Individualisierung zu treffen.

Anhand der in diesem Unterkapitel vorgegebene Usability Kriterien wird der MAP User Interface Agent im nächsten Unterkapitel bewertet.

Tab. 7.3: Beurteilung des MAP User Interface Agenten nach ISO 9241-10

Aufgabenan- gemessenheit	Der Dialog mit dem Benutzer unterstützt denselben angemessen, um die Aufgaben des Benutzers zu einer effizienten und effektiven Erledigung zu bringen. Die Angemessenheit erreicht dabei annähernd die im Projektziel geforderte delegative Güte und Assistenz.
Selbstbe- schreibungs- fähigkeit	Durch den Einsatz der konversationalen Metapher ist der Dialog in hohem Maße selbstbeschreibungsfähig. Jeder einzelne Dialogschritt wird dem Benutzer durch konversationale Rückmeldung visualisiert.
Steuerbarkeit	Durch die bewusste explizite Konversationsmodellierung wird dem Anwender jederzeit die Möglichkeit gegeben den Dialog zu beeinflussen. Dies wird tatsächlich durch die Offerte von Turn-takingpunkten bzw. durch Sequenzanzeigen weiter unterstützt.
Erwartungs- konformität	Durch den Einsatz der Konversation als steuernde Komponente der Dialogmodellierung kann eine hohe Erwartungskonformität bezüglich der generellen Anforderungen des Benutzers an Kommunikation erreicht werden. Dies ist ein Effekt, welcher sich aus der Nutzung von allgemeinen Konventionen bezüglich zwischenmenschlicher Kommunikation ergibt.
Fehlerrobust- heit	Durch den Einsatz von multimodalen Techniken kann eine Fehlerrobustheit gegenüber Fehleingaben des Benutzers erreicht werden, die deutlich höher ist als die Robustheit gängiger GUI.
Individuali- sierbarkeit	Die Individualisierbarkeit des User Interface Agenten ist sowohl in Bezug auf Arbeitsabläufe sowie Nutzerpräferenzen als auch in Bezug auf das individuelle Aussehen des konversationalen Charakters gegeben. Sie erfolgt über eine freie Wahl des Benutzers bezüglich der Animationsgrundlage des Charakters ebenso wie über Userprofiling und Benutzer-spezifische Definition von Agentenaufgaben.
Lernförder- lichkeit	Die Lernförderlichkeit wird im MAP System durch den Einsatz eines speziellen Hilfe-Agenten erreicht.

7.2.2 Usability Beurteilung

Anhand der verschiedenen Usability Prinzipien ist der MAP User Interface Agent, siehe Abbildung 7.6, einer kritischen Evaluierung unterzogen worden, vgl. [SH01]. Die Aufgaben des User Interface Agenten ist die Vermittlung zwischen dem Anwender und den Basisagenten des MAP Systems - und dies unabhängig von den Diensten, die die Basisagenten im Angebot führen.

Werden die Dialogprinzipien der ISO 9241 zugrundegelegt, so kann dem MAP User Interface

⁵ Vgl. Kapitel 2.2.

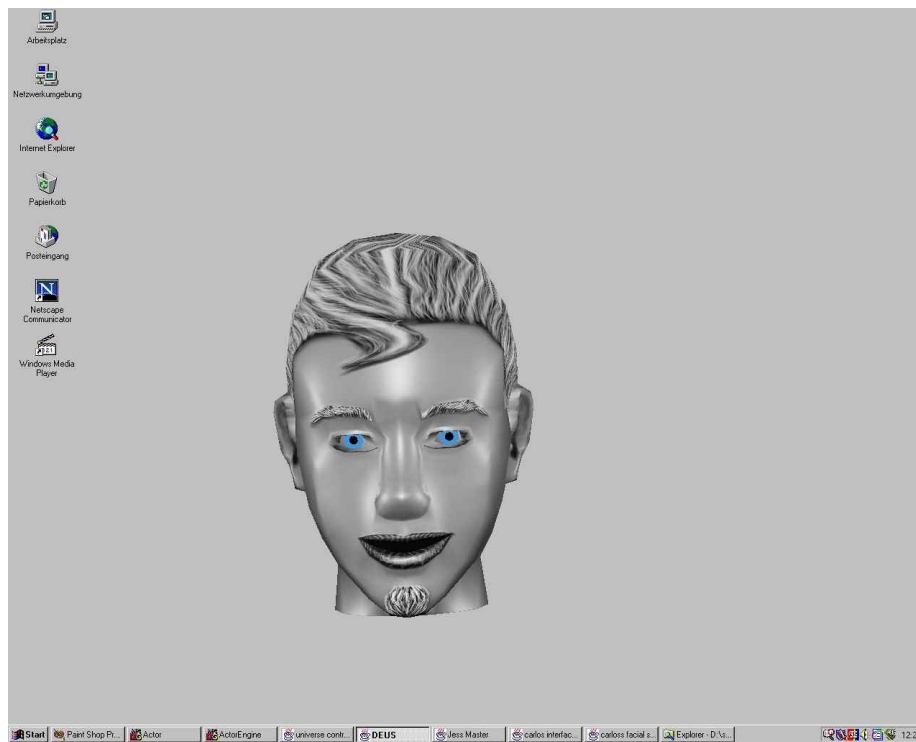


Fig. 7.6: MAP User Interface Agent Prototyp

Agenten eine Beurteilung gewährt werden, die dem klassischen GUI überlegen ist. Die Usability-Beurteilung wird in Tabelle 7.3 ausgedrückt.

Tab. 7.4: Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, Aufgabenteilung zwischen Mensch und Agent

Mensch	Agent
Formulierung der Suchanfrage	Gewährt Hilfestellung bei der Formulierung. Stellt Quellen für die Informationssuche. Darstellung und Vorbereitung der Treffermenge.
Bewertung der Treffermenge	Stellt Schlussfolgerungen über die Vorlieben des Anwenders.
Verfeinerung der Suchanfrage	Gibt Hilfestellung bei der Verfeinerung.

Die klassischen Usability Anforderungen der ISO 9241-10 entbinden nicht von der Prüfung der agentenspezifischen Usability Anforderungen, wie sie im Kapitel 7.2.1 definiert sind. Das MAP Agentensystem lässt sich wie in Tabelle 7.4, 7.5 und 7.6 beschrieben konkretisierend beurteilen, siehe [SH01].

Das Siemens Usability Labor [SH01] kommt zum Schluss, dass der MAP User Interface

Tab. 7.5: Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, Rollenverteilung zwischen Mensch und Agent

Mensch	Agent
<p>Teamleiter</p> <ul style="list-style-type: none"> • Auslösen und Kontrolle der Suchanfrage • Kontrolle Agentenprofil • Kontrolle eigenes Präferenzenprofil 	<p>Berater</p> <ul style="list-style-type: none"> • Hilfestellung bei der Formulierung/Verfeinerung von Suchanfragen • Vorschläge bez. Änderung der Präferenzen <p>Team Kollege</p> <ul style="list-style-type: none"> • Informationssuche in diversen Quellen • Angepasste Darstellung der Treffermenge

Tab. 7.6: Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, sonstige Anforderungen

Visualisierung	Die Visualisierung der Agententätigkeit im MAP System ist über den User Interface Agenten als Kommunikationsschnittstelle zwischen Benutzer und Basisagenten realisiert. Dies führt zu einer übersichtlichen Darstellung des Agentenangebots. Die Trennung der einzelnen Agenten erfolgt über die Trennung zwischen Aufgaben. Der User Interface Agent zeigt verschiedene Aufgaben explizit als Wechsel des konversationellen Diskurses mit dem Benutzer an.
Kontrolle	Der Anwender bleibt bezüglich des MAP Agentensystems immer der Herr der Lage. Über den User Interface Agenten ist jederzeit die Möglichkeit des Abbruchs von Aufträgen durch den Benutzer gegeben. Die grundlegenden Suchkriterien und Quellen sind für den Benutzer frei definierbar.
Vorhersagbarkeit	Der Benutzer ist in der Lage, die Suchstrategie der MAP Agenten zu hinterfragen und abzurufen. Aufgrund der in MAP verwendeten Security Konzepte ist der Agent jederzeit in der Lage, über die Vertrauenswürdigkeit seiner Informationen Auskunft zu geben.
Transparenz	Der Agent kann über den User Interface Agenten jederzeit den Aufenthalt und die Tätigkeit der einzelnen MAP Agenten überprüfen. Auf Nachfrage erfährt der Benutzer, welche Informationsquellen durchsucht wurden und ob und warum bestimmte Quellen nicht durchsucht wurden.

Agent aus ergonomischer Sicht sehr sinnvoll ist.

- Der MAP User Interface Agent stellt für den Benutzer die einzige konversationale Schnittstelle dar, was insgesamt eine Reduktion der Kommunikation mit dem Agentensystem bedeutet - der Benutzer muss nicht mit vielen Agenten gleichzeitig kommunizieren, sondern kann sich auf die Konversation mit dem User Interface Agenten konzentrieren.
- Der MAP User Interface Agent wird vom Benutzer als persönlicher Assistent wahrgenommen, der ihm bei der Verrichtung seiner Arbeiten durch die Vermittlung zum MAP Agentensystem unterstützt.
- Der MAP User Interface Agent kann multimodal angesprochen werden, was die Kommunikationsmöglichkeiten des Benutzers mit ihm erweitert und die Kommunikation insgesamt vereinfacht.
- Der MAP User Interface Agent beherrscht eine multimodale Ausgabe, er kommuniziert auf eine konversationale Art - unter Einbeziehung von typischen, menschlichen Verhaltensweisen, was dem Benutzer das Verständnis der Ausgaben des Systems erleichtert.
- Der MAP User Interface Agent kann parallel zu direkt manipulativen Interaktionen des Benutzers angesprochen werden, was eine zusätzliche Hilfestellung für den Benutzer darstellt.

Somit fällt die Beurteilung des MAP User Interface Agenten im Hinblick auf die Anforderungen Effektivität, Effizienz und Akzeptanz an das MAP Gesamtsystem positiv aus. Es konnte gezeigt werden, dass Delegation und Assistenz, die das MAP System über den User Interface Agenten gewährt, nutzbringend für den Benutzer einsetzbar ist. Tatsächlich sind die Voraussetzungen erfüllt, um die Effektivität der Arbeit des Benutzers zu steigern und die Effizienz der Arbeit zu erhöhen. Der einzelne User Interface Agent Repräsentant sichert dem MAP System eine höhere Akzeptanz durch den Benutzer als bei Verwendung von jeweils einem Repräsentanten pro MAP Agenten zu erzielen ist.

7.2.3 Anwendungstest

Die Beurteilung des MAP User Interface Agenten wird in einem anschließenden Anwendungs- und Feldtest validiert. In diesem Feldtest ist das MAP Gesamtsystem im Rahmen eines Nutzertests untersucht.

Folgende Funktionen sind im Testcluster getestet:

1. Terminmanagement unter der Berücksichtigung von Reisezeiten in Kombination mit Ressourcenmanagement.
2. Ressourcenmanagement.

3. Planung von Videokonferenzen.
4. Dialogführung mit dem Benutzer bei planungsbedingten Rückfragen über den Notifikationsdienst mit unterschiedlichen I/O Kanälen.

Das Szenario dieses Test wird Beispielhaft an Testszenario 1 dokumentiert, vgl. Fehr et al. [FKKM02]:

Innerhalb des Feldtests wird eine Terminvereinbarung zwischen maximal 4 Personen (sogenannten Akteuren⁶) durchgeführt.

Die einzelnen Aktionen eines Terminvereinbarungsprozesses definieren sich wie folgt⁷:

1. Der Initiator spezifiziert über die Anwenderschnittstelle eine Beschreibung für die zu planende Videokonferenz. Diese Beschreibung enthält Daten über den Ort, die früheste Start- bzw. die späteste Endzeit, die ungefähre Dauer des Treffens, die Teilnehmer etc..
2. Die Beschreibung wird an den MAP Termin Koordinations-Agenten versendet.
3. Der MAP Termin Koordinations-Agenten plant mit den Zeitplanungs-Agenten der Konferenzteilnehmer einen Termin.
4. Der MAP Termin Koordinations-Agent schickt die mit den Zeitplanungs-Agenten vorvereinbarten Termine an den MAP User Interface Agenten des Initiators.
5. Der MAP User Interface Agent des Initiators präsentiert dem Initiator die Auswahl der möglichen Termine.
6. Der Initiator wählt einen Termin aus der Vorschlagsliste aus.
7. Der MAP User Interface Agent schickt diesen Vorschlag an die MAP User Interface Agenten der Konferenzteilnehmer zur persönlichen Bestätigung.
8. Der Termin wird den einzelnen Konferenzteilnehmern zur persönlichen Bestätigung von ihren MAP User Interface Agenten vorgelegt.

Nach einer Einführung in den Testablauf und das grundlegende Systemverhalten wurde der eigentliche Test in Form einzelner Aufgabenstellungen durchgeführt⁸; jeder Testteilnehmer erhielt eine Liste mit Aufgaben, die im Feldtest zu lösen waren. Falls der Teilnehmer bei der Aufgabenlösung nicht weiter wußte bzw. Hilfe benötigte, wurde von dem Testleiter ein

⁶ Ein Akteur ist ein abstraktes Objekt (z.B. Person, Gerät, Softwaresystem) außerhalb eines Systems, das an Interaktion mit dem System beteiligt ist. In diesem Feldtest beschreiben Akteure Anwender des MAP-Systems. Bei den Akteuren werden nicht die beteiligten Personen unterschieden, sondern ihre Rollen (z.B. Initiator, Teilnehmer, etc.), die sie im Kontext des Anwendungsfalls annehmen. Dabei kann eine Person verschiedene Rollen annehmen.

⁷ Eine detaillierte Beschreibung der Testpläne findet sich in Fehr und TaQuang [FT02]

⁸ Eine detaillierte Beschreibung der Testabläufe findet sich in Fehr et al. [FKM02]

Hinweis zur Lösungsfindung gegeben. Um den Feldtest möglichst realistisch zu gestalten, wurden Terminkalender in Outlook erstellt, die bereits mit Terminen versehen waren, so dass eine Terminfindung nicht immer möglich war. Ebenso wurden die Teilnehmer in vier verschiedene Gruppen aufgeteilt. Die Gruppen A, B, C und D hatten verschiedene Aufgaben zu lösen. Jeder Teilnehmer war während des Feldtests abwechselnd Initiator und Akteur. Die Tests schlossen mit einem strukturierten Interview, in dem den Testteilnehmern die Gelegenheit gegeben wurde, ihre sowohl positiven als auch negativen Eindrücke von dem System zu artikulieren. Jeder Gesamttestlauf dauerte ca. zwei Stunden, es wurden vier Testläufe durchgeführt.

Tab. 7.7: Nutzerbeurteilung der Notifikation durch einen Avatar (Voting-Angaben in Prozentzahl der Benutzer)

	Sehr gut	Gut	Mittelmäßig	Eher schlecht	schlecht	Ent-haltungen
Sind die einzelnen Dialogschritte verständlich gestaltet?		75	25			
Wie beurteilen Sie die Bedienbarkeit des Dialogs insgesamt?		100				
Wie beurteilen Sie die Einteilung in Dialogeröffnung, -mitte und -ende?	25	75				25
Wie beurteilen Sie die Verwendung eines Avatars?	75	25				
Fühlen Sie sich durch den MAP Avatar beim Arbeitsvorgang unterstützt?	25	25	50			
Ist die Sprachausgabe hilfreich?	50	50				
Wie ist die Bedienbarkeit mittels Spracheingabe?	25	75				
Wie ist der Gewinn gegenüber den normalen Dialogboxen?	25	75				
Wie ist der Gewinn gegenüber dem MAP User Interface Agent-Multimodal-Browser?		50	25			25

Ein Schwerpunkt der Testdurchführung war die Beurteilung des konversationalen Avatars, siehe Tabelle 7.7. Diese fiel sehr positiv aus. Die Benutzer empfanden die Präsenz des

Avatars als eine gewisse soziale Nähe zum System. Die Beobachtung der Benutzer durch die Testleiter beim Test ergaben eine positive Grundstimmung, welche vor allem durch das konversationale und emotionale Auftreten des Avatars beeinflusst wurde. Die Benutzer zeigten sich Überrascht von der einfachen Bedienbarkeit mittels Sprachein- und ausgabe. Tatsächlich wurde der Avatar sofort als Ansprechpartner für das MAP angenommen. Des weiteren ist die Beurteilung des Dialogs eher mit gut als mit sehr gut bewertet. Es ist ein Gewinn gegenüber der normalen Dialogboxen und ein leichter Vorteil des personifizierten Systems (Avatar) gegenüber dem nicht personifizierten Multimodal-Browsers zu verzeichnen. Die Testergebnisse bestätigen somit die Ergebnisse der Usability Beurteilung⁹.

7.3 DIVA Evaluation

Die Evaluation des DIVA Systems erfolgt mittels (informatisch, psychologisch und filmwissenschaftlich begleiteten) Pre-Projekt-Studien, der Erarbeitung eines allgemeinen Szenarios sowie der Szenario Bewertung im Anschluss.

7.3.1 Studien

Im Hinblick auf die Konzeption einer API für nichtlineares Geschichtenerzählen und eines zugrunde liegenden Konzepts der interaktiven Narration von kontinuierlichen Medien wurden verschiedene Studien durchgeführt. Diese Studien befassen sich mit verschiedenen Aspekten des Nonlinear Storytelling wie:

- Möglichkeiten von Video on Demand.
- Möglichkeiten der Interaktiven Video Präsentation.
- Verbindungen zwischen Informatik und Filmwissenschaft in Digital Storytelling.
- Psychologische Grundlagen der Interaktion auf Video.

Von Braun und Blechschmitt, vgl. [BB97], wurde eine Studie über die Möglichkeiten des Business TV (BTV) in verteilten Unternehmungen angestellt. Zur Integration der BTV-Programme in die Arbeitsplatz-PC's sind verschiedene Konzepte entworfen und Realisierungsansätze bewertet. Zu diesem Zweck ist in [BB97] ein Katalog definiert, der Anforderungen an die BTV-PC-Integration nach verschiedenen Gesichtspunkten enthält. Die betrachteten Perspektiven schließen z. B. die Sicht des Anwenders, des Administrators und den Kostenaspekt ein. Mit Hilfe des Anforderungskatalogs wird ein Satz von Kriterien zur Bewertung von Lösungsansätzen abgeleitet. Ausgehend von einem Referenzmodell sind verschiedene Lösungen dargestellt und bewertet.

U.A. wurden dabei Ansätze zur Verteilung von Video auf Computern diskutiert und bewertet - der Ansatz einer digitalen Verteilung des TV Signals mittels eines Videoservers über

⁹ MAPUsabilityBeurteilung

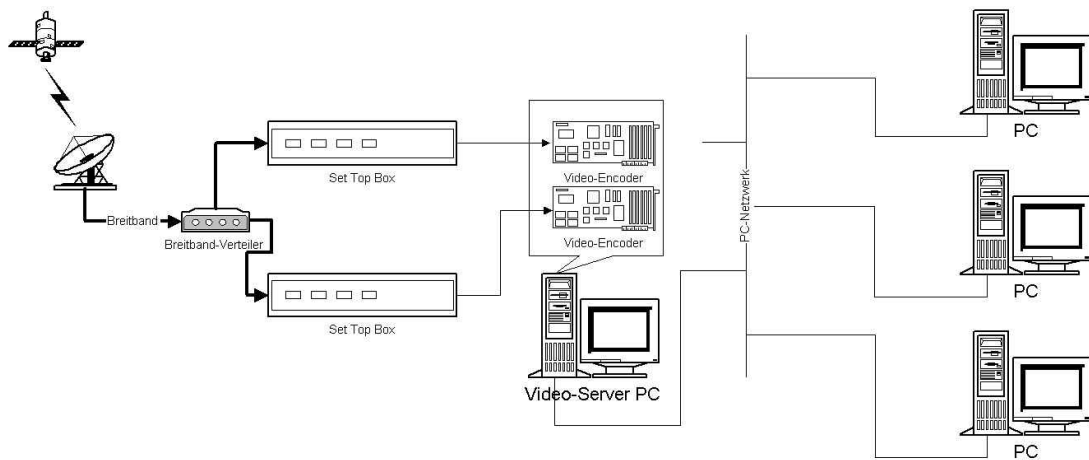


Fig. 7.7: Digitale Verteilung eines Fernsehsignals mittels eines Videoservers

ein Computer-Netzwerk, siehe Abbildung 7.7, fand dabei besondere Berücksichtigung und wurde als die sinnvollste Alternative des BTV im Hinblick auf interaktive Videopräsentationen gewertet. Die Alternative bietet folgende Eckpunkte:

- Jeder digitale Videostrom wird zum Transport über ein digitales Netzwerk vorbereitet - Der Video-Server PC (VS-PC) schickt den Videostrom in Transportform auf das PC-Netzwerk.
- Nutzung eines PC-Netzwerkes zur Verteilung des digital kodierten Fernsehsignals.
- Der Arbeitsplatz-PC empfängt den Videostrom in Form von Transportpaketen über das PC-Netz (Netzwerkkarte).
- Zur Dekodierung der Transportpakete und zum Aufbau eines Videostroms wird ein spezielles Videoclient-Programm benötigt.

Als Produktvorschlag wurde die digitale Video-Komplettlösungen RealPlayer genannt.

Tab. 7.8: Designkriterien für die Narration von nichtlinearen Geschichten

Kriterium	Bedeutung
Erzählstruktur	Eine Geschichte sollte mittels einer flexiblen Erzählstruktur automatisiert werden.
Dramaturgie	Die Dramaturgie der linearen Erzählung sollte trotz Interaktion des Zuschauers erhalten bleiben.
Interaktion	Die Interaktion sollte bis zu jenem Punkt gehen, bei dessen Erreichen das Erzählen einer Geschichte generell noch möglich ist

In einer Studie von Braun et al. [BSW99] wurden die möglichen synergetischen Effekte zwischen Informatik und Filmwissenschaft intensiv untersucht. Dabei wurde für die Ver-

bindung zwischen Film und Software in Industrie und Forschung die *interaktive Geschichte* als Symbiose zwischen der linearen Story der Filmindustrie und der offenen, interaktiven Anwendung mittels Software erkannt. Multimediale Anwendungen sind als das ideale Test- und Entwicklungsbett für diese Fusion der beiden Felder impliziert - eine Umgebung, in der sich beide Bereiche durch die Herausbildung von Schnittstellen treffen können. So wurde die Bestrebung der Filmwirtschaft zur Nutzung der neuen Medien ebenso erkannt wie die massive Verwertung von Wissen um filmische Dramaturgie in Computerspielen. Die Studie fordert die in Tabelle 7.8 notierten Designkriterien für die Narration von nichtlinearen Geschichten.

Weitere Studien¹⁰, wurden hinsichtlich der direkt manipulativen, intramedialen, graphischen Interaktion¹¹ auf Video getätigt. So wurden von Zahn et al. [ZSB99], unter anderem folgende Fragen gestellt:

- Wo sollte ein Hyperlink im Video platziert werden?
- Welche Videoobjekte sollten annotiert werden?
- Wie lange sollte ein Videoobjekt annotiert sein?
- Warum sollten entsprechende Videoobjekte annotiert sein?
- Welcher formale Aspekt des Videos impliziert die Annotation entsprechender Videoobjekte?
- Sollten Hyperlinks explizit (siehe Kapitel 4.3.2.1) markiert sein?
- Wie sollte zusätzliche Information zum Video präsentiert werden (im Hinblick auf GUI-Präsentation)

Diese Fragen wurden mittels einer explorativen Studie untersucht. Je drei Gruppen von Anwendern wurden anhand von verschiedenen Video-Clips bezüglich der oben notierten Punkte befragt. Die Anwendergruppen wurden wie folgt kategorisiert:

1. Experten über das Thema des Videos.
2. Experten bezüglich Multi Media.
3. Unerfahrene auf beiden oben genannten Gebieten.

Den drei Gruppen wurden nacheinander 3 Video-Clips vorgelegt. Die Video-Clips enthielten verschiedenartige Inhalte, die in Bild und Ton präsentiert wurden:

1. Der erste Clip enthielt eine Animation über die Struktur einer (biologischen) Zelle.
2. Der zweite Clip enthielt Fotos verschiedener biologischer Aspekte einer Zelle.

¹⁰ In Zusammenarbeit mit der Universität Tübingen

¹¹ Vgl. Kapitel 4.3.2.1.

3. Der dritte Clip zeigte Bilder über historische Aspekte der Biologie.

Die Clips wurden von den drei Gruppen anhand der oben notierten Fragen bewertet. Das Resultat dieser Bewertung wurde von Zahn [ZSB99] wie folgt notiert:

- Eine relativ hohe Anzahl von Hyperlinks (zumindest einer pro Sekunde) wurde für jeden Video-Clip angegeben. Die Anzahl unterschied sich je nach Inhalt und Ausführung des Videos. Das Platzieren der Hyperlinks unterschied sich nicht zwischen Gruppe 1 und 3, jedoch zwischen diesen beiden Gruppen und der Gruppe 2.
- Es existierten zwei Design-Strategien für die Annotation des Videos- eine bezüglich der inhaltlichen Aspekte, die zweite bezüglich formaler Aspekte des Videos.
- Die inhaltlichen Aspekte der Annotation wurden als medienunabhängig erkannt - sowohl graphische, graphisch-akustische als auch akustische Informationen wurden als Grund zur Annotation von Objekten angegeben.
- Als einzige Variable mit signifikantem Einfluss auf die Anzahl der gewünschten *graphischen* Hyperlinks wurde der akustische Part des Videos angegeben: Je weniger akustische Information zu der Graphik des Videos zu hören war, desto mehr graphische Annotationen wurden von den Testgruppen angefordert.
- Waren Informationen redundant, sowohl in der Graphik als auch in der Akustik des Videos angeboten, so forderten die Testgruppen eine Annotation sowohl der Graphik als auch der Akustik des Videos. Wurde Information alleine in der Akustik des Videos angeboten, so forderten die Testgruppen ebenfalls eine Annotation der Akustik des Videos.

Tab. 7.9: Designkriterien für die Annotation von Video, [ZSB99]

Das Video sollte redundante Information in Graphik und Akustik besitzen.
Jedes neue und relevante Informationsobjekt im Video sollte annotiert sein.
Nutze die Möglichkeit der Annotation von akustischen Inhalten.
Hyperlinks sollten explizit markiert sein.
Die Narration sollte stoppen, wenn eine Annotation ausgewählt ist und dem Benutzer zusätzliche Information angezeigt wird.
Zusätzliche Informationen sollten einen sehr engen inhaltlichen Bezug zum annotierten Videoobjekt besitzen.
Zusätzliche Information sollte eher statisch präsentiert werden.

Zahn trifft als Abschluss ihres Tests die Schlussfolgerung, dass die akustische Information des Videos einen starken Einfluss auf das Design von graphischen Annotationen des Videos hat. Sie fordert die Möglichkeit der Annotation des akustischen Teils des Videos als vom

Publikum geforderte Möglichkeit der Interaktion auf Videoinhalten. Neben diesen Forderungen stellt Zahn eine Liste von Design-Kriterien für Videoannotationen auf, die in Tabelle 7.9 notiert sind.

7.3.2 Szenario

Die Auswertung der verschiedenen Studien führt zu einem Szenario für interaktives, nicht-lineares Erzählen von Geschichten mittels Video. Folgende Aspekte finden dabei Berücksichtigung:

- Nichtlineare Narration auf einem Level, das dem Publikum die nötige Freiheit der Interaktion belässt, jedoch gleichzeitig zum Ziel einer Geschichte führt.
- Tatsächliche Möglichkeit des Publikums, konversational auf die Video Darbietung einzugehen.
- Explizite Annotation von Video, dies sowohl im graphischen als auch im akustischen Teil des Videos.
- Digitales Verteilen von Video, um eine umfassende Zugänglichkeit des Videomaterials zu erreichen.

Diese Aspekte wurden in der Konzeption für DIVA¹² umgesetzt. DIVA enthält folgende Komponenten:

- Interaktives Storytelling mittels einer Story Modellierung und einer Story Engine nach Propp¹³, diese wie durch Braun et al. [BSW99] gefordert sind.
- Konversationale Modellierung sowie konversationale Interaktionsmöglichkeit auf Video¹⁴, ebenso durch Braun et al. [BSW99] gefordert.
- Eine explizite, temporale und intramediale Möglichkeit der Annotation von Video, wie durch Zahn et al. [ZSB99] gefordert, als direkt manipulatives Interaktionskonzept¹⁵.
- Die Nutzung eines digitalen Videoservers¹⁶, wie er durch Braun und Blechschmitt, vgl. [BB97], vorgeschlagen wurde.

Ein Anwendungsszenario wurde in Form einer Web-basierten Video on Demand Möglichkeit, zugreifbar über das World Wide Web, umgesetzt, vgl. [Las00]:

¹² Vgl. Kapitel 6.3.

¹³ Vgl. Kapitel 6.3.2.2.

¹⁴ Vgl. Kapitel 6.3.3.

¹⁵ Vgl. Kapitel 6.3.2.1.

¹⁶ Vgl. Kapitel 6.3.2.

- Das Angebot von Videos über das Internet hat – neben der Tatsache, dass die Bestellung zu jeder beliebigen Tageszeit von zuhause aus erfolgen kann – den Vorteil, dass umfangreiche Zusatzinformationen leichter bereitgestellt werden können. Tools für die Suche von solchen Informationen arbeiten bei weitem effektiver, als ein realer Verkäufer es kann. Zur Erleichterung der Auswahl können Filmtrailer präsentiert werden, und der Benutzer hat zur Entscheidungsfindung soviel Zeit wie nötig.
- Um den gewohnten Kommunikationsformen des Benutzers Rechnung zu tragen, kommt ein konversationales Interface zum Einsatz. Als Ansprechpartner fungiert ein Avatar, der die gesamte Interaktion übernimmt. Genau abgestimmt auf die Zielgruppe bildet dieser die optimale Präsentationsweise. Dies wird vor allem dann deutlich, wenn das Benutzerverhalten analysiert wird, vgl. [Doy99]. Der Benutzer muss eine Auswahl treffen und benötigt dazu eventuell Beratung und zusätzliche Informationen. Daneben hat der Avatar eine soziale Rolle zu erfüllen. Er soll den User zu Aktionen anregen und die Aussagen der Trailer oder Videos durch sein Verhalten verstärken.
- Um die Zusatzinformationen zu den dazugehörigen Videos bereitzustellen, werden temporale Hyperlinks verwendet. Damit und durch den Dialog mit dem Avatar können die angebotenen Informationen exakt auf das jeweilige Video zugeschnitten werden. Der Rest der Website bleibt unabhängig vom eben gezeigten Trailer, da die filmspezifischen Daten nur im Videostream vorhanden sind. Damit ist ständige Aktualität gegeben.
- Die Benutzeroberfläche ist so gestaltet, dass fast alle Eingaben sowie ein Teil der Ausgabe mittels Konversation erfolgen können. Die Rolle des Ansprechpartners übernimmt ein Avatar. Die Navigation innerhalb der Website erfolgt über Spracheingabe.

Anhand dieses Szenarios wurde die Wirkung einer interaktiven Videonarration getestet: Einer Gruppe von Benutzern wurden zwei Video-Clips zur alternativen Auswahl gestellt:

- Technischer Video-Clip über den Aufbau eines Wintersportgerätes.
- Filmtrailer und Filmsequenzen.

Die Benutzer sahen sich in der Lage, sowohl die akustischen als auch graphischen Annotationen auszuführen, wenn sie direkte zusätzliche Information zu den Inhalten der Video-Clips benötigten. Die allgemeinen Fragen zu den Clips (beim Filmtrailer z. B. Fragen zu Regisseur oder Hauptdarstellern) wurden mittels des konversationalen Interfaces bearbeitet, die Benutzer fragten den Avatar, welcher ihnen - bei weiter laufendem Video - antwortete. Die Eingaben des Benutzers - sowohl die konversationalen als auch die direkt manipulativen - wurden zum selektieren weiterer Clips genutzt.

Als Resultat des Tests können die Thesen der Studien, siehe Kapitel 7.3.1, bezüglich der Interaktionsmöglichkeiten auf Video bestätigt werden. Sowohl die Nachfrage nach akustischen Annotationsmöglichkeiten von Video als auch die Nachfrage nach einer nichtlinearen Narrationsmöglichkeit mit konversationaler Interaktionsmöglichkeit wurden als tatsächlich nutzbar belegt.

7.4 Resultate und Diskussion

In diesem Abschnitt werden die aus den Anwendungen gewonnenen Resultate fixiert und das Konzept der interaktiven Narration von kontinuierlichen Medien bewertet. Dabei lassen sich unterschiedliche Bewertungskriterien aufstellen. So wird überprüft, wie umfassend das Konzept tatsächlich alle Komponenten, die zur nichtlinearen Narration von Video benötigt werden, realisiert: sind nichtlineare Narration, Interaktion mittels Konversation und direkter Manipulation auf Video, sowie Videoservice tatsächlich gegeben und sind diese als nonlinear Storytelling API tatsächlich für verschiedene Applikationen nutzbar? Die Resultate bemessen das Erreichen der in Kapitel 3.4 aufgestellten Anforderungen an nichtlineares Video sowie die Güte des erreichten Zieles. Darüber hinaus erfolgt eine Konzept-Beurteilung anhand der Resultate hinsichtlich des Gesamtzusammenhangs mit dem Problemkreis, welcher in der Einleitung dieser Arbeit motiviert wurde. Dieses sind die Tendenzen zur Informationsplattform für Jederman, zur Unterstützung des Benutzers durch Assistenz und Delegation, zum Einbeziehen von sozialen Komponenten im Umgang mit Computern, zur konversationalen Interaktion mit Computern und zur Konvergenz zwischen Software-Industrie und Film- und Medienwirtschaft durch interaktive mediale Erlebnisse mittels Computer.

Die vollzogene Umsetzung des Konzepts mit einer integrierten Bewertung der Güte der Ansätze wird im Folgenden diskutiert.

Konzept konversationale Interaktion Die konversationalen Möglichkeiten der Interaktion sind sowohl in der Evaluation des Projektes EMBASSI als auch in MAP als vorhanden und anwendbar getestet wurden. Die explizite symbolische Modellierung von Konversation wurde im MAP Projekt umgesetzt und bewährt sich dort als zentrale Komponente der Kommunikation zwischen System und Benutzer. Es zeigt sich, dass die konversationale Modellierung ein eigenständiger, von den Inhalten der Konversation weitgehend unabhängiger Teil des User Interfaces ist. Darüber hinaus kann sogar festgestellt werden, dass die konversationale Modellierung als zusätzliche Komponente des User Interface einen Mehrwert für den Benutzer beinhaltet.

Neben der bestätigten Umsetzung des Konzeptes zeigen die Anwendungen in EMBASSI, MAP und auch DIVA, dass die API zum Ansprechen der konversationalen Modellierung in verschiedene Applikationen eingebunden werden kann und somit die Anwendungsprogrammierung erheblich erleichtert. Neben der konversationalen Modellierung wird ein Direktzugriff auf einen Charakter zur Verfügung gestellt.

Konzept direkt manipulative Interaktion Die direkt manipulative Interaktionsmöglichkeit, die Applikations- und Medien-spezifisch für den Bereich Video konzipiert und realisiert ist, orientiert sich gänzlich an den in Kapitel 7.3.1 notierten Designkriterien für Annotation von Video. So wird damit ein Ansatz validiert, der sowohl die temporalen und spatialen Eigenschaften des Mediums Video als auch die Ansprüche des Nutzers an eine solche Interaktionsmöglichkeit berücksichtigt:

- Der Benutzer erfährt die Annotation des Videos explizit. Das bedeutet, dass er tatsächlich jederzeit nachvollziehen kann, wo und wann eine direkt manipulative

Interaktionsmöglichkeit in Bezug auf die Videoinhalte gegeben ist.

- Der Benutzer erfährt die temporale Struktur der Videoannotation: Sowohl in Bezug auf die Graphik als auch in Bezug auf die Akustik hat der Benutzer jederzeit Gewissheit über die temporale Ausdehnung der Annotation.
- Die Videoannotation erfolgt intramedial - eine Umorientierung zwischen Präsentationsmedium und Annotationsmedium durch den Benutzer muss nicht erfolgen.

Anhand der Eigenschaften kann verifiziert werden, dass der in Kapitel 5.2.1 spezifizierte Ansatz komplett umgesetzt und für den Benutzer nachvollziehbar und effizient ist.

Die konzeptionelle Umsetzung ist - konzeptbedingt, da Applikationsspezifisch gehandhabt - über ein SMIL API für den entsprechenden Videoserver durch den Anwendungsprogrammierer nutzbar¹⁷.

Konzept Nichtlineare Videopräsentation Die applikationsspezifische Präsentation von Medien ist mit einem nichtlinearen Videoservice, der die grundlegenden asynchronen Interaktionsmöglichkeiten zur Verfügung stellt, realisiert. Dabei ist die in Kapitel 7.3.1 notierte Anregung bezüglich des Videoservers aufgegriffen und realisiert. Die Resultate bezüglich der DIVA Studie lassen den Schluss zu, dass das Konzept der applikationsspezifischen Medienpräsentation erfüllt ist.

Das Ansteuern des Servers erfolgt für den Anwendungsprogrammierer über das SMIL Protokoll¹⁸.

Konzept nichtlineare Narration Dieses Konzept ist nach den Anforderungen laut Kapitel 7.3.1 umgesetzt. Die nichtlineare Narration erfolgt nach einem Ansatz von Vladimir Propp, der ein Schema von Geschichten erstellt und eine Möglichkeit der interaktiven Erzählweise präsentiert. Dieses Schema ist in einem speziellen morphologischen Geschichtenmodell umgesetzt und durch eine allgemeine, morphologische Story Engine ausführbar. Morphologisch indizierte Videosequenzen werden interaktiv dem Benutzer dargeboten, der trotz größtmöglicher interaktiver Freiheit auf das Ziel (Ende) der Geschichte zwingend zusteuert. Als Resultat lässt sich schließen, dass tatsächlich eine nichtlineare Narration der Geschichte erfolgt und das Konzept somit in diesem Bereich umgesetzt ist.

Die API der nichtlinearen Storynarration ermöglicht dem Anwendungsprogrammierer ein Ansteuern der Story Engine sowie die Definition weiterer Story Modelle.

Das Zusammenspiel der Komponenten, wie es zum Beispiel mittels der DIVA Studie gezeigt wurde, kann als effizienter, holistischer Ansatz des nonlinear Storytelling mit einem besonderen Bezug zur Narration der Geschichte (Erzähler) und einer Beeinflussung der Narration durch direkte Manipulation (Handlung der Zuhörer) und Konversation (Ansprechen des Erzählers) gewertet werden, wie er in der Einleitung gefordert ist, siehe Abbildung 1.4.

¹⁷ Vgl. Kapitel 3.3.1.

¹⁸ Vgl. Kapitel 3.3.1.

Insbesondere wird hiermit die Vollständigkeit des Ansatzes bezüglich interaktiver Narration mittels Video gezeigt.

Neben den Schnittstellen für Applikationen und Anwendungsprogrammierung kann das gesamte Konzept durch entsprechende Autoren zu spezifischen Zwecken einem einfachen Redesign unterzogen werden. Dies ist im Projekt MAP gezeigt worden, das Design von verschiedenen Konversationsmodellen zulässt. Aufgrund der abstrakten, symbolischen Modellierung sowohl des konversationalen Ansatzes als auch des nonlinearen Storytelling Ansatzes kann der Autor die jeweiligen Modelle durch ein explizites Verändern der jeweiligen Modellstruktur beeinflussen. So können bezüglich des konversationalen Modells zusätzliche konversationale Aspekte sowie Konversationssituation-verändernde Regeln in das System eingepflegt werden. Das Story Modell kann mittels Einführung von weiteren morphologischen Funktionen sowie weiteren Beziehungen zwischen den Funktionen erweitert und modifiziert werden. Dies geschieht auf einer symbolischen, abstrakten Stufe von Wissensmodellierung, die eine Konzentration auf die Thematik durch den Autoren nicht durch Details der Programmierung untergräbt.

Bezüglich des Problemkreises, der in der Einleitung spezifiziert wurde, kann die Umsetzung des Assistenz- und Delegationsgedankens mittels des konversationalen Teilkonzepts¹⁹ als vollzogen angesehen werden. Hierbei bieten sowohl die Projekte EMBASSI und MAP allgemein nutzbare Informationsplattformen an, welche die Weiterentwicklung des Mediums Computer vom Einzelplatzrechner zur Kommunikationsumgebung für Jedermann belegen. Die soziale Komponente der Interaktion zwischen Menschen wird dabei auf die Interaktion mit den Medien übertragen. Durch die nachgewiesene Umsetzung der nichtlinearen, interaktiven Narrationsmöglichkeit mittels der Story Engine kann ebenfalls die Konvergenz der klassischen narrativen Medien wie Film und Buch mit der Informatik in einen weiteren Schritt vorangetrieben werden. Tatsächlich kann mit der expliziten Formulierung von Story Modellen der narrative Teil der Ästhetik von Geschichten automatisiert werden. Eine besondere Erleichterung für den Anwendungsprogrammierer ist die Definition von APIs zu allen zuvor genannten Aspekten, die erst die Nutzung der entsprechenden Techniken - sei es die Nutzung von Teilkonzepten oder die Nutzung der kompletten nichtlinearen, interaktiven Narrationsmethodik - durch einen breiten Kreis von Anwendungen gewährleistet. Insofern können die in Kapitel 1.1 aufgestellten Arbeitshypothesen als erfüllt angesehen werden.

Durch die Bereitstellung von diesen API ist erst die breite Nutzung der Methodiken des nichtlinearen Storytelling in Industrie und Wirtschaft ermöglicht. Dies erfolgt - wie in diesem Kapitel belegt - zum Nutzen des Anwenders, dem durch die bereitgestellten Strukturen ein interessantes, immersives, dramatisches und soziales Erlebnis mit dem Computer - ob an seinem Arbeitsplatz oder im heimischen Umfeld - angeboten wird.

¹⁹ Insbesondere im Hinblick auf die Resultate des EMBASSI Projekts und des MAP Projektes, siehe Kapitel 7.1 und 7.2.

8. ZUSAMMENFASSUNG UND AUSBLICK

Dieses Kapitel dient der Zusammenfassung der durchgeführten Arbeiten und erzielten Ergebnisse. Dabei werden die Ergebnisse insbesondere hinsichtlich der Fragestellungen von Graphisch Interaktiven Systemen und Digital Storytelling diskutiert. Die Anwendungsgebiete der Thematik und Konzepte dieser Arbeit werden, ausgehend vom Anwendungsbeispiel, erweitert. Schließlich werden im Rahmen eines Ausblicks weiterführende Arbeiten benannt.

Die vorliegende Arbeit hat ein Konzept zur Narration und Interaktion für nichtlineare Erzählformen auf kontinuierlichen Medien am Beispiel Digitales Video - kurz interaktive Narration von kontinuierlichen Medien - aufgestellt. Dieses Konzept wurde an Hand der Weiterentwicklung von Computern von Einplatz-Systemen zu Kommunikationsplattformen, des Wechsels von Anwendungsmethoden - weg von der Werkzeugmethode, hin zu Assistenz und Delegation - der Forderung nach neuen, sozialen und konversationalen Umgangsformen mit dem Computer sowie der synergetischen Beziehung zwischen Film- und Medienwirtschaft und Software Industrie motiviert. Die Motivation legte eine Videospezifische Betrachtungsweise des Problems nahe und führte zur Aufstellung von zwei Arbeitshypothesen:

1. Interaktivität kann als dramaturgisches und immersions-steigerndes Mittel vorteilhaft zur Narration von kontinuierlichen Medien, zum Beispiel Video, eingesetzt werden.
2. Konversation kann als Metapher der Kommunikation mit Computersystemen vorteilhaft zur Assistenz und Delegation durch den Benutzer eingesetzt werden.

Es wurden systematisch die Anforderungen des Anwendungsfeldes untersucht und eine Anzahl von Anforderungen zur Lösung der Thematik aufgestellt. Als erster Schritt wurde die derzeitigen Anwendungsfelder von Videopräsentationen und Avataren sowie die kombinierte Anwendung von beiden Bereichen notiert, siehe Kapitel 2. Des weiteren wurde eine Analyse von Nichtlinearen Storytelling Systemen vorgenommen, deren Grundlagen diskutiert und aktuelle Ansätze und Konzepte vorgestellt, siehe Kapitel 3. Mögliche Interaktionsformen auf kontinuierlichen Medien wurden, ausgehend von deren Grundlagen über die Vorstellung aktueller Konzepte, in Kapitel 4 diskutiert. Wird das Anwendungsfeld den traditionellen Gebieten der Computergraphik gegenüber gestellt, so ist vor allem die Forderung nach einer interaktiv beeinflussbaren Struktur der Narration von Geschichten - daher ein interaktiver Erzähler von Geschichten - hervorzuheben. Dieser Erzähler soll auf zumindest zwei Arten beeinflussbar sein - dieses durch die konversationale Interaktion, welche dem Publikum zur

natürlichen, sozialen Kommunikation mit dem Erzähler geboten wird sowie durch die Beeinflussung der Geschichte - deren Medien und Strukturen - und durch direkte Handlungen des Publikums, daher durch einen direkt manipulativen Ansatz. Gefordert wird ein abstraktes Konzept der interaktiven Narration am Beispiel von Video, welches jedoch durch die Angabe von konkreten Komponenten - sowohl deren Struktur als auch die zugrunde liegenden Modelle - und entsprechenden Möglichkeiten, diese Komponenten mittels einer API - zusammenfassend Nonlinear Storytelling API genannt - ganz konkret ausformuliert und ansteuerbar sein soll.

Die Lösungsidee besteht aus einer Trennung von inhaltlichen, story-strukturellen und konversationalen Aspekten der Narration, siehe Kapitel 5, vgl. [Bra02d]. Die Idee zeichnet sich dadurch aus, dass die Erzählstruktur (Story) als auch die Erzählweise (diegetische (Konversation) und mimetische (mediale) Narration) separat von den Inhalten entwickelt werden kann. Insbesondere kann sie, im Gegensatz zu sonstigen Systemen, symbolisch notiert werden und unterstützt damit die Einbeziehung von Experten auf dem jeweiligen Gebiet. Zu diesem Zweck werden Methoden der Künstlichen Intelligenz - insbesondere der Wissensbasierten Systeme - für die Modellierung des Expertenwissens bezüglich der beiden Domänen verwendet. Die Lösungsidee ist generisch¹ einsetzbar - mittels des Lösungsansatzes können verschiedenste Modelle der Narration und der Konversation beschrieben werden, vgl. [Bra02a].

Die konversationale Modellierung ist direkt von Beobachtungen zwischenmenschlichen, konversationalen Verhaltens abgeleitet. Sie verfolgt damit einen Top-Down Ansatz, welcher vom generellen Konversationsverhalten des Menschen ausgeht, die auf die Maschine zu übertragenden Anteile kritisch auswählt, modelliert und erst in einem weiteren Schritt modal umsetzt, vgl. [Bra02f]. Für die modale Umsetzung ist eine Charakter Engine angegeben, die amodale Konversationsanweisungen interpretiert und in konkretes Verhalten eines virtuellen Actors (Avatars) umsetzt, vgl. [BS01b].

Die Modellierung der interaktiven Narration der Story folgt einem semiotischen Ansatz, welcher sowohl eine eindeutige Klassifizierung von Geschichten mittels Morphologien als auch eine abstrakt gehaltene Vorgehensweise zur nichtlinearen Narration von Geschichten nutzt. Diese Vorgehensweise wurde in dieser Arbeit konkretisiert und erweitert sowie auf die Narration von kontinuierlichen Medien angepasst, vgl. [BG01].

Aus der Erkenntnis der Medienabhängigkeit von direkter Manipulation ist der direkt manipulative Anteil der Interaktion mit Geschichten medienspezifisch gelöst. Die Lösung zeichnet sich dabei insbesondere durch einen direkt manipulativen Ansatz der synchronen Interaktion auf Video aus. Dieser Ansatz, der sowohl die temporalen Eigenschaften des Mediums berücksichtigt als auch der medialen Ausprägung - Video als kombiniertes Medium aus Bild und Ton - Rechnung trägt, wird explizite temporale Videoannotation genannt [Bra01a].

Der Lösungsidee folgend wurden Grundlagen der verschiedenen Aspekte der Lösung analysiert, aus dieser Analyse wurde ein Konzept - das Lösungskonzept der interaktiven Narration von kontinuierlichen Medien - zur Verwirklichung der Idee abgeleitet. Das Konzept kann-

¹ Eine Anpassung auf weitere Applikationsbereiche, zum Beispiel Computer Games, ist denkbar, vgl. Braun [Bra02c].

te - in Gänze oder in Teilen - als Basis verschiedener Systeme eingebracht werden. Durch die Separation der Lösung in einzelne Komponenten konnte außerdem ein flexibler Einsatz auch von Teilkomponenten des Systems verifiziert werden.

Zur Validierung des Konzepts sind drei verschiedene Systeme vorgestellt worden, welche das Konzept - gänzlich oder in Teilen - implementieren, siehe Kapitel 6. So wurden die Konzepte in die Projekte **EMBASSI** und **MAP** eingebracht sowie als DIVA Studie realisiert. Durch die Anwendung des Konzepts in diesen Projekten und mittels der Projekt-bezogenen Validierungsergebnisse konnte gezeigt werden, dass das Konzept - sowohl in Teilen als auch als Ganzes - nutzbringend und im Sinne der Arbeitshypothesen eingesetzt wird, siehe Kapitel 7. So wurde die konversationale Modellierung als probates Mittel zur Kommunikation mit dem Benutzer bei Assistenz und Delegationswünschen durch denselben im Projekt **MAP** bewertet. Durch die Zentrierung der Agenten-Benutzer Kommunikation auf einen User Interface Agenten, der die konversationale Ausgestaltung des Dialoges mit dem Benutzer übernimmt und somit den Basisagenten des MAP Systems von dieser Aufgabe befreit, wurde ein soziales Interface geschaffen, das dem Benutzer als persönlicher Assistent dient. Die nichtlineare Story Modellierung wurde als zielführendes und dramaturgisch einsetzbares, interaktives Schema notiert und der Einsatz im Bereich Videonarration und Präsentation diskutiert. Die Arbeitshypothesen werden somit als erfüllt notiert. Im Hinblick auf die in Kapitel 2 notierten Anforderungen an Digital Storytelling bietet der in dieser Arbeit vorgestellte Ansatz und API eine integrierte multimediale und multimodale Erfahrung, welche durch die morphologische Story Engine orchestriert wird.

Die Autorenunterstützung wird in Arbeiten am ZGDV weiter verfolgt. So wurde bisher ein Konzept des Szenario Based Design, vgl. [BLSG00], erstellt, das eine Hilfe zur strukturierten Entwicklung von Software am Beispiel von Geschichten beschreibt. Weitere Arbeiten, welche die Erstellung von Storys mittels der in dieser Arbeit vorgestellten Story Modellierung durch graphische Entwicklungstools unterstützen, werden zur Zeit konzipiert.

Die Arbeiten an den verschiedenen Komponenten werden zur Zeit auf weitere Bereiche des Digital Storytelling ausgedehnt. So kommen die konversationale Modellierung und die nichtlineare Narration in dem BMBF geförderten Projekt **Geist** (ein mobiles AR-Informationssystem zum Erleben historischer Zusammenhänge im urbanen Umfeld mit Digital Storytelling, vgl. Braun et al. [BDSG02], Braun [Bra03b] and Braun and Schneider [BS03]) und in dem ZGDV Eigenforschungsprojekt **Information zum Anfassen (IZA)**, vgl. Braun [Bra02a], zum Einsatz. Diese Projekte übertragen die grundlegenden Konzepte dieser Arbeit (Konversation und nichtlineare Narration) auf die Gebiete Virtual Reality (VR) und Augmented Reality (AR).

Das Projekt Geist konzeptioniert und implementiert eine AR basierte, interaktive Geschichte. Die Idee ist das historisch fundierte Erlebnis der vergangenen Geschichte einer Stadt - dies vor Ort in der Stadt, unter Einbeziehung von historischen Persönlichkeiten und Charakteren, historischer Gebäude und Plätze sowie historischer Handlung. Der Benutzer von Geist soll in eine immersive, dramatische Story hineingezogen werden, in der er selbst die Story interaktiv beeinflusst und damit intensiv erlebt, sich mit Charakteren - Geistern - auseinandersetzt, siehe Abbildung 8.1, und dabei die Historie der Stadt durch die Story



Fig. 8.1: Geist - Die Benutzerin interagiert mit einem virtuellen Avatar - einem 'geisterhaften' Bewohner des Heidelberger Schloss. Links: Benutzerin mit AR Equipment; Rechts: AR Sicht auf Katharina, einen Geist.

vermittelt bekommt. Die in dieser Arbeit vorgestellte Story Engine dient als Basis der Narration der Geschichte. Da die Szenerie in einem AR Umfeld stattfindet, welche nicht mit Video arbeitet, sind keine vorgegebenen Szenen (Video-Clips) vorhanden. So steuert die Story Engine eine entsprechende Szenen Engine an, welche die Szenen nach den entsprechend ausgewählten morphologischen Funktionen der Story Engine umsetzt. Auf Ebene der Szenen Engine werden wiederum die konversationalen Modellierungen der vorliegenden Arbeit zur Kommunikation mit den Benutzern des Systems genutzt.

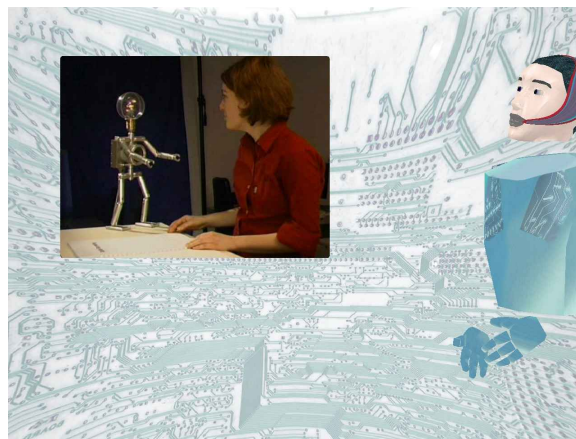


Fig. 8.2: IzA, virtueller Charakter interagierend mit Benutzer und physischem Charakter

Im Projekt IZA wird ein selbsterklärendes Kiosksystem konzipiert und implementiert. Der Kiosk nutzt verschiedenste reale Properties, die in Kombination mit Virtual Reality (VR) Darstellungen eine interessante, dramaturgisch auf einen Kurzaufenthalt des Benutzers am Kiosk abgestimmte Präsentation über ein (festzulegendes) Thema veranschaulichen. Die Interaktion des Benutzers geschieht dabei über eine soziale, konversationale Komponente

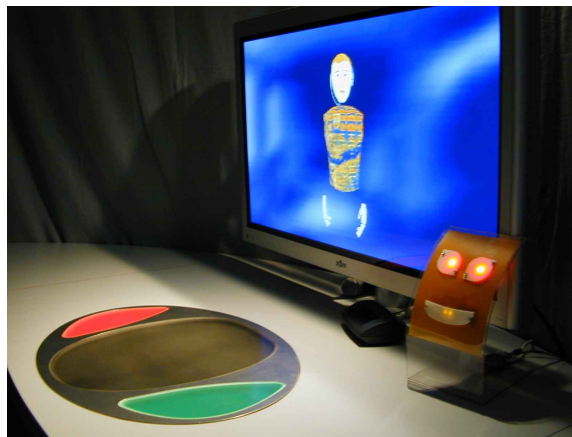


Fig. 8.3: IzA, virtueller Charakter, physischer Charakter und Interaktionsflächen

in Form eines Virtual Characters, siehe Abbildung 8.2 und Abbildung 8.3. Dieses Kiosksystem nutzt die Story Engine sowie ein modifiziertes Story Modell, um den beschriebenen Kurzaufenthalt des Benutzers in eine dramatische, interaktive Struktur abzubilden. Die konversationale Modellierung erfolgt mit den konversationalen Komponenten der hier vorgestellten Arbeit.

Neben den oben beschriebenen Projekten zur Weiterführung der hier vorgestellten Arbeiten in den Bereichen VR und AR geschieht eine Erweiterung des Konzepts der konversationalen Modellierung. Das in der vorliegenden Arbeit zur Kommunikation zwischen Mensch und Maschine eingesetzte Konzept wird erweitert auf die Abbildung von konversationalem Gruppenverhalten. Dieses Verhalten ist unter anderem von André [AR00] als Verhalten zwischen Avataren und Benutzern beschrieben worden. Diese Arbeiten werden um konversationale Aspekte von Gruppen erweitert. Der konversationale Aspekt zeigt - innerhalb einer Benutzergruppe einer Kommunikationsplattform - den jeweiligen generellen konversationalen und sozialen Zustand der Gruppe an. Die Modellierung entspricht daher einem oder mehreren Metawesen, was die Gruppe der Benutzer repräsentiert, siehe Braun und Rieger [BR01] und Braun et al. [BRF01].

Die grundlegenden Storytelling-Techniken dieser Arbeit wurden um die Konzepte des Narrative Conflict und der Suspense Progression erweitert, siehe Braun und Rieger [BR03], um eine angepasste Ausdrucksweise eines Virtual Characters zu erreichen.

Hier wurde im Bereich der Gebärdensprache-Präsentation eine Plattform geschaffen, welche es erlaubt, automatisiert realistische Gebärdensprache darzustellen, siehe Rieger und Braun [RB03]. So zeigt Abbildung 8.4 eine freundliche Aufforderung, sich zu nähern, während Abbildung 8.5 eine ärgerliche Aufforderung, sich zu nähern, zeigt. Deutlich sind die Unterschiede in der Ausdrucksweise der Gestik und der Mimik zu erkennen.

Das in dieser Arbeit vorgestellte Konzept kann, neben den konkreten Projekten und Entwicklungen, welche oben angegeben sind, in genereller Hinsicht auf die Entwicklung von Graphisch Interaktiven Systemen und der Informatik gewertet werden. So ist eine Erweite-

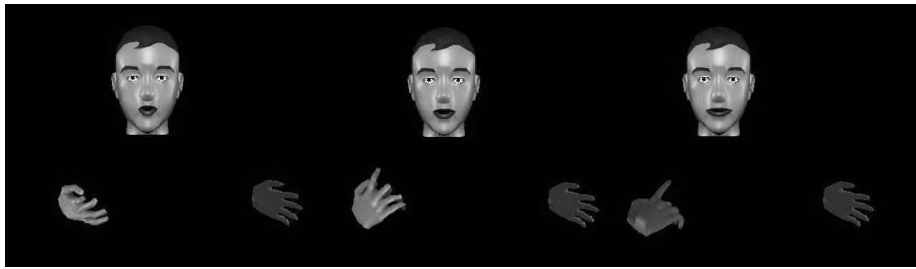


Fig. 8.4: Gebärdensprache und Mimik in einer freundlichen Ausdrucksweise

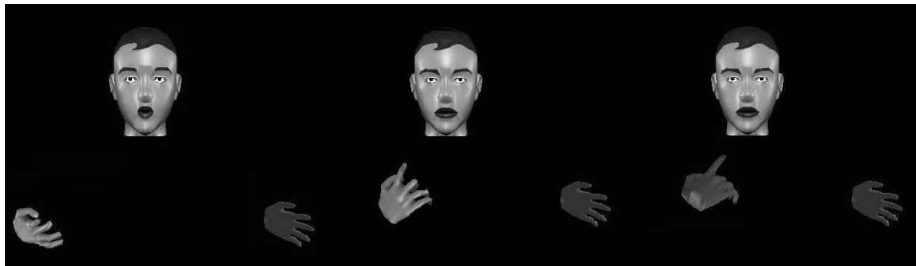


Fig. 8.5: Gebärdensprache und Mimik in einer ärgerlichen Ausdrucksweise

rung des Gebietes um intelligente Strukturen notwendig, die nicht nur eine Assistenz und Delegation sowie eine Mensch-ähnliche Kommunikation erlauben, sondern dem Benutzer den Umgang mit der Maschine durch VR und AR Techniken tatsächlich transparent gestalten, so dass der Übergang vom Einen zum Anderen schließlich verwischt. Hierzu ist weitere Arbeit bezüglich der graphisch-interaktiven Abbildung der Denk- und Erzählweise von Menschen notwendig, dies sowohl im Bezug zur nichtlinearen Narration als auch zur konversationalen Interaktion, welche sich als Komponenten der Computer Graphik gerade etablieren.

Der Ausblick zeigt, dass die vorliegende Arbeit einen Startpunkt für weiterführende Arbeiten auf dem Gebiet der Konversation und der interaktiven Narration bietet. Die in dieser Arbeit vorgestellte API für nonlinear Storytelling bildet dabei einen generellen Rahmen für weiterführende Arbeiten. Die Verknüpfung von Konversation und Narration bleibt bei weiterführenden Arbeiten erhalten - dies schon Aufgrund der Eigenschaften von Narration, welche sowohl diegetische als auch mimetische Aspekte umfasst und somit eine kombinierte Betrachtung und Diskussion der Gebiete nahelegt, wie sie in dieser Arbeit vorgenommen wurde.

ANHANG

LITERATURVERZEICHNIS

- [ACL01] Fipa acl message structure specification. <http://www.fipa.org/specs/>, 2001.
- [act00] Activeworlds, 15. April 2000. URL, <http://www.activeworlds.com>.
- [AES⁺00] Maria Aretoulaki, Christian Elting, Michael Strube, Stefan Rapp, Georg Michelitsch, Claudia Settele, P. Hilt, and Vitor Sa. Embassi - spezifikation a-phh, dialogmanager und polymodaler input/output manager. Arbeitspapier - spezifikation aphh, Embassi Konsortium, 2000.
- [AGGS96] Marc Alexa, Norbert Gerfelder, Paul Grimm, and Christian Seiler. Avwod - concept and realization for internet-based media integration. In P. Hoschka, editor, *Proceedings of the Real Time Multimedia and the World Wide Web. W3C Workshop 1996*, pages pp 14/1–14/7. W3C, 1996.
- [Ale94] David Alexander. *Gene Roddenberry: Die autorisierte Biographie*. Wilhelm Hyne Verlag GmbH, München, Germany, 1994. ISBN: 3-453-12794-3.
- [AM99] Mark Alexa and Wolfgang Müller. The morphing space. In Watzlav Skala, editor, *Proceedings of 7-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Pilsen, Czech Republic, 1999.
- [ana00] Ananova, die virtuelle nachrichtensprecherin, 28. April 2000. URL, <http://www.ananova.com>.
- [apo00] Cyberflash, 6. April 2000. URL, <http://www.apollinaris.de>.
- [AR00] Elisabeth André and Thomas Rist. Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. In *Proceedings of the Second International Conference on Intelligent User Interfaces*, pages 1–8, 2000.
- [ARM97] Elisabeth André, Thomas Rist, and Jochen Müller. Webpersona: A life-like presentation agent for the world-wide web, workshop "intelligent education systems on the world wide web". In *AI-ED 97, 8th World Conference on Artificial Intelligence in Education*, Kobe, Japan, 1997.
- [Aro92] Barry Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 1992.

- [ask00] Welcome to desktop assistant, 27. März 2000. URL, <http://www.askltd.com/askjava/Intro.htm>.
- [Ast98] Peter Astheimer. Acoustiv simulation for visualisation and virtual reality. Eurographics State of the Art Report, 1998.
- [Atc01] Denise Atchley. The digital storytelling conference & festival. <http://www.dstory.com/>, 2001.
- [Aus00] Australian broadcasting corporation, abc news online, 10. Mai 2000. URL, <http://www.abc.net.au/news/>.
- [ava00a] Avatare von avatars, 3. April 2000. URL, <http://www.avatara.com/avatars2/avatars1.html>.
- [ava00b] Avatarecasino, 4. April 2000. URL, <http://www.avatarecasino.com>.
- [Ayl99] Ruth Aylett. Narrative in virtual environments - towards emergent narrative. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. AAAI, 1999.
- [Bac00] Kent Bach. Routledge encyclopedia of philosophy, entry speech acts, 2000. URL, <http://userwww.sfsu.edu/~kbach/spchacts.html>.
- [Bah90] Hans Paul Bahrtdt. *Schlüsselbegriffe der Soziologie. Eine Einführung mit Lehrbeispielen*. C.H. Beck, München, Germany, 4 edition, 1990.
- [Bar00] Mark Barret. Graphics: The language of interactive storytelling. *Computer Graphics*, 34(3):7–9, August 2000.
- [BARM00] Norbert Braun, Marc Alexa, Thomas Rieger, and Wolfgang Mueller. Ein flexibles präsentationssystem für animierte user-interface-agenten und avatare. In Ulrike Spierling, editor, *Digital Storytelling Tagungsband*, pages 29–35, Darmstadt, Germany, 2000. Gesellschaft für Informatik und Zentrum für Graphische Datenverarbeitung, Fraunhofer IRB Verlag.
- [Bat93] Joseph Bates. *Virtual Realities: Anthology of Industry and Culture*, chapter The Nature of Characters in Interactive Worlds and The Oz Project. 1993.
- [Bau95] Penny Bauersfeld. User interface challenges of media design. In *3. Internationale Multimedia Konferenz*, San Francisco, USA, November 1995.
- [BB97] Norbert Braun and Eric Blechschmitt. Business tv. Fraunhofer Institut Computer Graphic FIGD 97v001-FIGD, 1997.
- [BB01] Eric Blechschmitt and Axel Brümmer. Map ze 2.1.22: Spezifikation der dialogbeschreibungssprache. Technical report, Fraunhofer Institut Graphische Datenverarbeitung, Darmstadt, 2001.

-
- [BBL⁺01] Eric Blechschmitt, Axel Brümmer, Volker Luckas, Brit Hockauf, and Jörg Sahm. Officeplus. abschlussbericht 2000. Report, Fraunhofer IGD, Darmstadt, 2001.
- [BC97] Stephen A. Brewster and Catherine V. Clarke. The design and evaluation of a sonically-enhanced tool palette. In *Proceedings of the International Conference on Auditory Displays ICAD 1997*, 1997.
- [BD98a] Norbert Braun and Ralf Doerner. Sonic hyperlinks: Hypermedia methodologies applied to audio for www-based teaching applications. In *Proceedings of the World Conference of the WWW and Internet, WebNet'98*, 1998.
- [BD98b] Norbert Braun and Ralf Doerner. Using sonic hyperlinks for web-tv. In *Proceedings of the International Conference on Auditory Displays ICAD 1998*, 1998.
- [BD99] Norbert Braun and Ralf Doerner. Temporal hypermedia for multimedia applications in the world wide web. In *Proceedings of the 3rd International Conference on Computational Intelligence and Multimedia Applications IC-CIMA '99*, 1999.
- [BDSG02] Norbert Braun, Johanna Dechau, Oliver Schneider, and Stefan Goebel. The story as contextual basis of the collaborative group experience. ACM Conference on Collaborative Virtual Environments 2002, Workshop on Storytelling in Collaborative Virtual Environments, Bonn, Germany, 2002, 2002.
- [BEG⁺00] Kai Bücher, Christian Elting, Rolf Gaidzik, Michael Hellenschmidt, Detlef Jantz, Merdad Jalali, Nicole Krämer, Wolfgang Meyer-Schmidt, and Julia Nitschke. Bedienassistentz für embassi - komponenten. EMBASSI - Arbeitspapier - A-PHH phhINP049GRU04, Embassi Konsortium, Darmstadt, 2000.
- [Ber00] Arthur Asa Berger. Arthurs computer - (geschichten) abenteuer. *Television, Internationales Zentralinstitut für das Jugend- und Bildungsfernsehen*, 13(1), 2000.
- [BF99] Selmer Bringsjord and David A. Ferrucci. Artificial intelligence and literary creativity: Inside the mind of brutus, a storytelling machine. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. AAAI, Lawrence Erlbaum, Mahwah, 1999.
- [BF00] Norbert Braun and Matthias Finke. Interaction of video on demand systems with human-like avatars and hypermedia. In *IDMS2000, 7th Workshop on Interactive Distributed Multimedia Systems and Telecommunications Services*, Enschede, Niederlande, 2000. angenommen, noch nicht veröffentlicht.
- [BG95] Ludger Bölke and Peter Gorny. Direkte manipulation von akustischen objekten durch blinde rechnerbenutzer. In H.-D. Böcker, editor, *Software-*

- Ergonomie 95 - Fachtagung der German Chapter of the ACM und der Gesellschaft für Informatik*, pages 93–105. ACM, Gesellschaft für Informatik, B.G. Teubner, 1995.
- [BG01] Norbert Braun and Dieter Grasbon. A morphological approach to interactive storytelling. In Monika Fleischmann and Wolfgang Strauss, editors, *Proceedings of the Conference on artistic, cultural and scientific aspects of experimental media spaces*, Bonn, Germany, September 2001. CAST.
- [BGS99] Norbert Braun, Manfred Gaida, and Ulrike Spierling. Holistic approach for the use of multi media and multi modality: Digital storytelling. In *Proceedings of the World Conference of the WWW and Internet, WebNet'99*, 1999.
- [BGS01a] Norbert Braun, Manfred Gaida, and Oliver Schneider. Map ze 2.1.12: Analyse von verhaltensmodellen in szenarios in hinsicht auf konversationsähnliche dialogführung mit ui-agenten. Technical report, ZGDV, Darmstadt, 2001.
- [BGS01b] Norbert Braun, Manfred Gaida, and Oliver Schneider. Map ze 2.3.10a: Dialogbeispiele für die konversationsschnittstelle. Technical report, ZGDV, Darmstadt, 2001.
- [Bib03] Wolfgang Bibel. *Lehren vom Leben*. Deutscher Universitäts-Verlag, Wiesbaden, Germany, 1 edition, 2003.
- [Ble01] Eric Blechschmitt. Map ze 2.4.4a: Dialogmanagement. Technical report, Fraunhofer Institut Graphische Datenverarbeitung, Darmstadt, 2001.
- [BLK⁺97] Gene Ball, Dan Ling, David Kurlander, John Miller, David Pugh, Tim Skelly, Andy Stankosky, David Thiel, Van Dantzich Maarten, and Trace Wax. Life-like computer characters: the persona project at microsoft research. In J.M. Bradshaw, editor, *Software Agents*, Cambridge, MA, 1997. MIT, MIP Press.
- [BLSG00] Norbert Braun, Ulrike Lohde, Ulrike Spierling, and Manfred Gaida. Scenario-based design: Geschichten als hilfsmittel beim erstellen von software. In Ulrike Spierling, editor, *Digital Storytelling Tagungsband*, pages 29–35, Darmstadt, Germany, 2000. Gesellschaft für Informatik und Zentrum für Graphische Datenverarbeitung, Fraunhofer IRB Verlag.
- [BNG89] Mel H. Brown, Stephen L. Newsome, and Ephraim P. Glinert. An experiment into the use of auditory cues to reduce visual overload. In *Proceedings of the CHI*, pages 339–346, USA, 1989. ACM, ACM Press, Addison-Wesley.
- [Bol80] Richard A. Bolt. Put that there: voice and gesture at the graphics interface. *Computer Graphics*, 14(3):262–270, 1980.
- [Bor85] David Bordwell. *Narration in the Fiction Film*. Madison University of Wisconsin Press, London, UK, 1985.

-
- [BPSMM00] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, and Eve Maler. Extensible markup language (xml) 1.0 (second edition). W3C Recommendation, October 2000.
- [BR01] Norbert Braun and Thomas Rieger. Group conversation within a internet tv community through a lifelike avatar. In *Proceedings of the International Conference on Media Futures*, Italie, 2001. AEI.
- [BR03] Norbert Braun and Thomas Rieger. Expressiveness generation for virtual characters, based on the principles of suspense progression and narrative conflict. In *Proceedings of the International Workshop on Mobile Computing, IMC 2003*, Rostock, Germany, June 2003.
- [Bra69] Ivo Braak. *Poetic in Stichworten: Literaturwissenschaftliche Grundbegriffe*. Verlag Ferdinand Hirt, Würzburg, Germany, 1969.
- [Bra99] Norbert Braun. Conversational video interaction. In *Proceedings of the World Conference of the WWW and Internet, WebNet'99*, 1999.
- [Bra00] Norbert Braun. Konversationale benutzerschnittstellen im projekt map. Jahresbericht ZGDV, 2000.
- [Bra01a] Norbert Braun. Interaction approach for digital video based storytelling. In *Proceedings of the 9-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2001*. EUROGRAPHICS, Computer Graphics Society and IFIP, 2001.
- [Bra01b] Norbert Braun. Modeling of conversational user interface. International Conference on Text Speech and Dialogue, 2001. Tutorial.
- [Bra01c] Norbert Braun. Nonlinear storytelling api. ZGDV Interner Bericht, Dezember 2001.
- [Bra02a] Norbert Braun. Automated narration - the path to interactive storytelling. Fraunhofer Publica- Datenbank, 2002. Workshop on Narrative and Interactive Learning Environments, Edinburgh, Scotland.
- [Bra02b] Norbert Braun. Menschähnliche konversation - eine diegetische narrationskomponente für delegation und assistenz. In *Proceedings of the VDI/VDE - GMA Fachtagung Useware 2002*, Darmstadt, Germany, 2002. VDI/VDE.
- [Bra02c] Norbert Braun. Narrative semiotics for computer games. Workshop Challenge of Computer Games, Lodz, Poland, 2002, 2002.
- [Bra02d] Norbert Braun. Programmierte narration und konversation, eingesetzt zum interaktiven erzählen von geschichten. In *Proceedings of the Mensch & Computer 2002, Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten*, Hamburg, Germany, 2002. GI & German Chapter of ACM.

- [Bra02e] Norbert Braun. Storytelling & conversation to improve the fun factor in software applications. Fraunhofer Publica- Datenbank, 2002. CHI 2002 Conference on Human Factors in Computing Systems, Workshop Funologie, Minneapolis, Minnesota, SIGCHI, ACM, USA.
- [Bra02f] Norbert Braun. Symbolic conversation modeling used as abstract part of the user interface. In Václav Skala, editor, *Proceedings of the 10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, Februar 2002. EUROGRAPHICS and IFIP working group 5.10.
- [Bra03a] Norbert Braun. *Funology: From Usability to Enjoyment*, chapter Storytelling & conversation to improve the fun factor in software applications. Kluwer Academic Publishers, Dordrecht, Netherlands, April 2003.
- [Bra03b] Norbert Braun. Storytelling in collaborative augmented reality environments. In Václav Skala, editor, *Proceedings of the 11-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, Februar 2003. EUROGRAPHICS and IFIP working group 5.10.
- [Bre63] Berthold Brecht. *Das epische Theater. Schriften zum Theater 3*. Frankfurt am Main, Germany, 1963.
- [BRF01] Norbert Braun, Thomas Rieger, and Matthias Finke. Community tv: An approach to interaction for groups and single users on internet video. In *Proceedings of International Conference on Multimedia, Internet, Video Technologies*, Malta, September 2001. WSES.
- [BS89] Joseph Bates and Sean Smith. Towards a theory of narrative for interactive fiction. Technical Report CMU-CS-89-121, School of Computer Science, Carnegie Mellon University, Pittsburgh PA., February 1989.
- [BS01a] Eric Blechschmitt and Christoph Strödecke. Map ze 2.4.3a: Dialoginterpretation. Technical report, Fraunhofer Institut Graphische Datenverarbeitung, Darmstadt, 2001.
- [BS01b] Norbert Braun and Oliver Schneider. Conversation modeling as an abstract user interface component. In *Proceedings of GI Workshop Synergien zwischen Virtueller Realität und Computerspielen: Anforderungen, Design, Technologien*, Wien, Österreich, September 2001. GI und OCG, Fachausschuss 4.7 der GI e.V. und Fachbereich Medizinische Informatik der GMDS e.V.
- [BS03] Norbert Braun and Oliver Schneider. Suspenseful user experiences in collaborative virtual spaces, enabled by interactive narration. In *Proceedings of the HCII International 2003, 10th International Conference on Human - Computer Interaction*, Crete, Greece, June 2003.

-
- [BSH02] Norbert Braun, Oliver Schneider, and Gregor Habinger. Literary analytical discussion of digital storytelling and its relation to automated narration. HCI Europe '2002, Workshop Understanding User Experience: Literary Analysis meets HCI, London, UK, 2002.
- [BSR⁺01] Eric Blechschmitt, Jörg Sahm, Volker Roth, Saied Tazari, Christian Klee-grewe, Stefan Müller, Dung TaQuang, Michael Fehr, and Zdenek Mikovec. Ze 3.2.1: Spezifikation der systemschnittstellen. Technical report, Fraunhofer Institut Graphische Datenverarbeitung, Darmstadt, 2001.
- [BSW99] Norbert Braun, Ulrike Spierling, and Claudia Weber. Neue entwicklungs-methoden für digital storytelling: Ein vergleich der etablierten bereiche der film-produktion und der software-entwicklung und ihre relevanz für die moderne multimedia-anwendungsentwicklung. ZGDV, 1999.
- [Bux93] William A. S. Buxton. Hci and the inadequacies of direct manipulation sy-stems. *SIGCHI Bulletin*, (25(1)):21–22, 1993.
- [Bux95] William A. S. Buxton. *Readings in Human Computer Interaction: Toward the Year 2000*, chapter Speech, Language & Audition. Morgan Kaufmann Publishers, San Francisco, 1995.
- [Bux98] William A. S. Buxton. An introduction to this special issue on interaction in 3d graphics. *Computer Graphics: The SIGGRAPH Quarterly*, 32(4):43–44, 1998.
- [BWK92] Joseph Bates, Peter Weyhrauch, and Margaret Thomas Kelso. Dramatic pre-sence. Technical Report CMU-CS-92-195, School of Computer Science, Car-negie Mellon University, Pittsburgh PA., December 1992.
- [Cam99] Joseph Campbell. *Der Heros in tausend Gestalten*. 1999.
- [CBB⁺99] Justine Cassell, Timothy W. Bickmore, Mark Billingham, L. Campbell, K. Chang, Hannes Hogni Vilhjalmsson, and H. Yan. Embodiment in con-versational interfaces: Rea. In *Proccedings of the CHI '99*, pages 520–527, USA, 1999. ACM, ACM Press, Adison -Wesley.
- [ccc00] cccc: Software specification assistant, 27. März 2000. URL, <http://www.cccc.com/ssa>.
- [CFC⁺97] Colleen E. Crangle, Lawrence M. Fagan, Robert W. Carlson, Mark S. Erl-baum, David D. Sherertz, and Mark S. Tuttle. Collaborative conversational interfaces. SMI Report SMI-97-0672, Stanford University, Stanford University, USA, 1997.
- [Cla68] Arthur C. Clarc. *2001 - Odyssee im Weltraum. Der Roman zum Film*. Wil-helm Hyne Verlag GmbH, München, Germany, 1968. ISBN: 3-453-19438-1.

- [cli00] Climark: Advisors assistant, 27. März 2000. URL, <http://www.climark.com/>.
- [Con00] Contact Consortium. Theu, virtual university, 24. April 2000. URL, <http://www.ccon.org/theu/index.html>.
- [Cor01] NTT Software Corporation. The future is here - be there! <http://www.nttsoft.com/>, 2001.
- [cou00] Courttv: Videos von gerichtsverhandlungen, 10. Mai 2000. URL, <http://www.court tv.com/video/>.
- [Cra82] Chris Crawford. *The Art of Computer Game Desig*, chapter A Taxonomy of Computer Games. <http://www.vancouver.wsu.edu/fac/peabody/gamebook/Coverpage.html>, 1982.
- [Cra99] Chris Crawford. Assumptions underlying the erasmatron interactive storytelling engine. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. ACM, 1999.
- [Cra00] Chris Crawford. Erasmus game engine homepage, 2000. URL, <http://www.erasmatazz.com>.
- [Cra01] Chris Crawford. Interactive storytelling. *Computers and Graphics Journal*, 2001.
- [CTCL96] Zhigang Chen, See-Mong Tan, Roy H. Campbell, and Yongcheng Li. Real time video and audio in the world wide web. In *Proceedings of the Fourth International World Wide Web Conference*, 1996.
- [CTP98] Justin Cassel, Obed Torres, and Scott Prevost. *Machine Conversations*, chapter Turn taking versus Discourse Structure: how best to model multimodal conversation. Kluwer, The Hague, 1998.
- [cyb00a] Cyberflash, 4. April 2000. URL, <http://www.cplus.fr/html/cyberflash/98-99/index.html>.
- [cyb00b] Cybertown, 3. April 2000. URL, <http://www.cybertown.com>.
- [Dat] www.kiotodate.com.
- [det00] Distance education and technology, university of british columbia, 10. Mai 2000. URL, <http://videos.cstudies.ubc.ca/>.
- [DGS01] Ralf Doerner, Paul Grimm, and Christian Sailer. Etoile - an environment for team, organizational and individual learning in emergencies. *Computer Graphik topics*, 3, 2001.
- [Doy99] Patrick Doyle. When is a communicative agent a good idea? workshop on communicative agents. In *Third International Conference on Autonomous Agents*, Seattle, USA, Mai 1999.

- [DPP⁺96] Thierry Dutoit, Vincent Pagel, Nicolas Pierret, Van Der Vreken Olivier, and Francois Bataille. Mbrola project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *Proceedings of the ICSLP*, 1996.
- [Dra] www.dramatica.com.
- [e-c00] E-cyas, 3. April 2000. URL, <http://www.e-cyas.com>.
- [Eis67] Sergei Eisenstein. *Film Form and The Film Sense*. Meridian Books, Cleveland, 1967.
- [EK00] José L. Encarnação and Thomas Kirste. Beyond the desktop: Natural interaction and intelligent assistance for the everyday life. In *Proceedings of Paderborner Podium 3: Alltag der Zukunft - Informationstechnik verändert unser Leben.*, pages 38–57, Paderborn, Schöningh, 2000. Heinz Nixdorf MuseumsForum (HNF). ISBN: 3-506-76233-8.
- [Ell95] Holger Ellermann. *Das Drehbuchschreiben als Handwerk*. Coppingrave, 1995.
- [EM00] Christian Elting and Georg Michelitsch. Embassi - pmo-regelwerk. Arbeitspapier - spezifikation pmo regelwerk, Embassi Konsortium, 2000.
- [EMR⁺00] Christian Elting, Georg Michelitsch, Thomas Rieger, Norbert Braun, Wolfgang Müller, and Maria Aretoulaki. Embassi - feinspezifikation pmo. Arbeitspapier - pmo-feinspezifikation-aphh, Embassi Konsortium, Darmstadt, 2000.
- [Enc97] José L. Encarnação. *Frontiers of Human-Centred Computing, Online Communities and Virtual Environments*, chapter Challenges and Frontiers of Computer Graphics A Vision for an Applied Research Agenda, pages 367–393. Springer-Verlag, Berlin, Germany, 1997. ISBN 1-85233-238-7.
- [Enc99a] José L. Encarnação. Digital storytelling. In Ralf Reichwald and Manfred Lang, editors, *Tagungsband Anwenderfreundliche Kommunikationssysteme*, pages 323–359, Heidelberg, Germany, Juni 1999. Münchner Kreis, Uebernationaler Vereinigung für Kommunikationsforschung, Forum Telekommunikation. ISBN: 3-7785-3937-X.
- [Enc99b] José L. Encarnação. *Schlüsseltechnologien für intelligente Produkte und Dienstleistungen*, chapter Zukunftsweisende F&E-Trends und Anwendungen fuer neue Medien in der Informations- und Kommunikationstechnologie, pages 15–41. VDI F+E Forum, Mannheim, Germany, 1999.
- [ES99] José L. Encarnação and Ulrike Spierling. Digital storytelling - die nächste api-generation für multimedia-anwendungen. SAVE Management-Club 1/99, Februar 1999.
- [Fie87] Syd Field. *Drehbuchschreiben für Film und Fernsehen*. 1987.

- [FKKM02] Michael Fehr, Feride Karabiyik, Christian Kleegrewe, and Stefan Müller. Systemtest basisagenten und anwendungsintegration für feldtest. MAP Interner Bericht Nr. map-capc010-020627-v01, 2002.
- [FKM02] Michael Fehr, Feride Karabiyik, and Stefan Müller. Auswertung feldtest cluster 3 und 7. MAP Interner Bericht Nr. map-capc012-020819-v04, 2002.
- [FKPT98] Ludwig Fahrmeir, Rita Künstler, Iris Pigeot, and Gerhard Tutz. *Statistik - Der Weg zur Datenanalyse*, volume 2. Springer Verlag, 1998. ISBN 3-540-65053-9.
- [FT02] Michael Fehr and Gong TaQuang. Erstellung testpläne für feldtest. MAP Interner Bericht Nr. map-capc007-020521-v01, 2002.
- [Gav89] Bill Gaver. The sonic finder: An interface that uses auditory icons. In *Human Computer interaction Vol.4*, USA, 1989. ACM.
- [Gle97] Uli Gleich. *New Horizons in Media Psychology*, chapter Parasocial Interaction with People on the Screen. 1997.
- [Goe94] Phil Goetz. *Interactive Fiction and Computers*, chapter Interactive Fantasy 1, pages 98 – 115. Crashing Boar Books, London, 1994.
- [Goe60] Johann Wolfgang von Goethe. *Notes and Essays on the West-Eastern Divan (1819)*, in *Werke 2 of Goethes Werke*, volume 14. Germany, 1948-60.
- [Gra00] Alexander Graham. Client/server videosystem mit integrierter hypermedia- und konversationsinteraktion. Diplomarbeit, Fachhochschule Darmstadt, Technische Universität Darmstadt, 2000.
- [Gri75] Paul H. Grice. *Syntax and Semantics: Speech Acts*, chapter Logic and Conversation. Academic Press, New York, 1975.
- [Han97] Chris Hand. A survey of 3d interaction techniques. *Computer Graphics forum*, 16(5):269–281, 1997.
- [Han00] Sven Handrick. Arbeitswissenschaftliche kriterien zu map. Zwischenergebnis map-tudr002, MAP Konsortium, 2000.
- [Har85] Mary Dee Harris. *Introduction to Natural Language Processing*. Reston Publishing Co. Inc, Reston, VA, USA, 1985. ISBN: 0-8359-3254-0.
- [Hei91] Edmund Heinen. *Industriebetriebslehre: Entscheidungen im Industriebetrieb*. Gabler, Wiesbaden, 9 edition, 1991.
- [Hen97] Norbert Henze. *Stochastik für Einsteiger*. Vieweg, Wiesbaden, Germany, 1997. ISBN 3-528-06894-9.
- [HG97] Brad Hefta-Gaub. The realmedia platform architectural overview. Real Networks Conference, 1997.

-
- [HK01] Thorsten Herfet and Thomas Kirste. Embassi - elektronische multimediale bedien- und service-assistenz. EMBASSI Whitepaper (Draft2) (Rev. 4) et-cINP003ZGD04, Embassi Konsortium, Darmstadt, 2001.
- [Hö98] Kristina Höök. Steps to take before intelligent user interfaces become real. In *Proceedings of 4. Designing and Evaluating Intelligent User Interfaces*, San Francisco, USA, 1998. ACM, ACM.
- [HS68] Roy Huss and Normann Silverstein. *The Film Experience*. Dell Publishing, New York, 1968.
- [Hy193] Jeremy Hylton. The complete works of william shakespeare. <http://the-tech.mit.edu/Shakespeare/>, 1993.
- [Inc98] OrangeIgloo Inc. Electric games review: Blade runner by westwood studios. <http://www.electricgames.com>, 1998.
- [ISO98] Ergonomische anforderungen für bürotätigkeiten mit bildschirmgeräten - teil 10: Grundsätze der dialoggestaltung (din iso 9241-10:1996). Beuth Verlag GmbH, 1998.
- [JND⁺00] Cathryn Johns, David Nunez, Marc Daya, Duncan Sellars, Juan Casanueva, and Edwin Blake. The interaction between individuals' immersive tendencies and the sensation of presence in a virtual environment. In J.D. Mulder and R. van Liere, editors, *Proceedings of 6th Eurographics Workshop on Virtual Environments*, Amsterdam, Netherlands, 2000. EUROGRAPHICS, Springer Verlag.
- [Joy87] Michael Joyce. Afternoon, a story. Cambridge, M.A.: The Eastgate Press, 1987. Computer Disk.
- [KA98] James E. Katz and Philip Aspden. Social and public policy internet research: Goals and achievements. Vortrag an der University of Michigan School of Information, 2. Februar 1998.
- [Kaa98] Eija Kaasinen. Usability issues in agent applications: What should the designer be aware of? In K-J. Räihä, editor, *Advanced course on human-computer interaction. Proceedings of ACHCI 1998*, Finland, 1998. ACHCI, University of Tampere.
- [kel00] Keller products inc., usa, 10. Mai 2000. URL, <http://www.kellerproducts.com/m25-50.html>.
- [KK97] Maz Kessler and Robby Kilgore. Bringing an agent to life. In *Animated Interface Agents: Making them intelligent*, Nagoya, Japan, August 1997.
- [Kle00] Marion Klein. Mate dialogue annotation guidelines. Deliverable - d2.1, MATE Multilevel Annotation, Tools Engineering, Telematics Project LE4-8370, 2000.

- [KN01] Nicole C. Krämer and Julia Nitschke. Ausgabemodalitäten im vergleich: Verändern sie das eingabeverhalten von benutzern? In *Proceedings of 4. Berliner Werkstatt Mensch-Maschine-Systeme, Thema: Bedienen und Verstehen*, Berlin, Deutschland, Juni 2001. Technischen Universitaet Berlin.
- [Koe99] Rob Koenen. Mpeg-4 multimedia for our time. *IEEE Spectrum*, 36(2):26–33, Februar 1999.
- [Koe01] Rob Koenen. Overview of the mpeg-4 standard. V.18 – Singapore Version, March 2001.
- [KP01] MAP Konsortium and MAP Projektbüro. Multimedia-arbeitsplatz der zukunft: Managementzusammenfassung. Technical report, Fraunhofer Institut Graphische Datenverarbeitung, Darmstadt, 2001.
- [KS97] Minoru Kobayashi and Chris Schmandt. Dynamic soundscape: mapping time to space for audio browsing. In *Electronic Publications of the CHI 97, USA*, 1997. ACM.
- [Las00] Jutta Laserer. Internet video-mate: Konversationales interface zur präsentation interaktiver videos. Diplomarbeit, Fachhochschule Hagenberg, Fachhochschule Hagenberg, Österreich, 2000.
- [Lau90] Brenda Laurel. *The Art of Human Computer Interface Design*, volume 25, chapter Interface Agents: Metaphors with Character. 1990.
- [Lau93] Brenda Laurel. *Computers as Theatre*. Addison-Wesley, New York, USA, 1993.
- [LB97] A. Bryan Loyall and Joseph Bates. Personality-rich believable agents that use language. In *Proceedings of the Agents '97*, Marina del Rey, CA, USA, 1997. ACM, ACM Press.
- [LBSA91] Brenda Laurel, Joseph Bates, Rachel Strickland, and Don Abbe. Interface and narrative arts: Contributions from narrative, drama and film. In *Proceedings of CHI '91*. ACM, ACM Press, Addison-Wesley, 1991.
- [LF94] Yanis Labrou and Tim Finin. A semantics approach for kqml – a general purpose communication language for software agents. In *Proceedings of the third International Conference on Information and Knowledge Management*, 1994.
- [LMW97] Bride Linane-Mallon and Brian Webb. Evaluating narrative in multimedia. In *Proceedings of the 4th International Eurographics Workshop*, pages 83–98. Eurographics, DSV-IS'97, 1997.
- [Lor92] Otto Lorenz. *Kleines Lexicon literarischer Grundbegriffe*. Wilhelm Fink Verlag, München, Germany, 1992.

-
- [Mar96] Robert Martel. A distributed network approach and evolution to html for new media. In P. Hoschka, editor, *Proceedings of the Real Time Multimedia and the World Wide Web. W3C Workshop 1996*. W3C, 1996.
- [mar00a] Marcmoini: Clean-install assistant, 5. Mai 2000. URL, <http://www.marcmoini.com/C-IAssistant.html>.
- [MAR⁺00b] Wolfgang Müller, Marc Alexa, Thomas Rieger, Norbert Braun, and Ulrike Spierling. Embassi - spezifikation avatar-plattform. Arbeitspapier - avatar-plattform-feinspezifikation-aphh, Embassi Konsortium, Darmstadt, 2000.
- [Mat97] Michael Mateas. An oz-centric review of interactive drama and believable agents. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh PA., June 1997.
- [Mau01] Fritz Mauthner. Die metaphor. <http://www.mauthner-gesellschaft.de/>, 2001.
- [Mol91] Sammy Molcho. *Körpersprache im Beruf*. Goldman, München, 1991. ISBN 3442163269.
- [MS99] Michael Mateas and Phoebe Sengers. Introduction to the narrative intelligence symposium. In *Proceedings of the AAAI 1999 Fall Symposium on Narrative Intelligence*. American Association on Artificial Intelligence, 1999.
- [MS00] Michael Mateas and Andrew Stern. Towards integrating plot and character for interactive drama. In K. Dautenhahn, editor, *Proceedings of the 2000 Fall Symposium: Socially Intelligent Agents: The Human in the Loop*, pages 113 – 118. AAAI Press, Menlo Park, CA, 2000.
- [Mur98] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. The MIT Press, MA, 1998.
- [NB97] Tsukasa Noma and Norman I. Badler. A virtual human presenter. In *Proceedings of the IJCAI '97 workshop on Animated Interface Agents - Making them Intelligent*, San Francisco, USA, August 1997. Morgan-Kaufmann Publishers.
- [Nel92] Theodor Holm Nelson. *Literary Machines 93.1*. Mindful Press, Sausalito, CA, USA, 1992.
- [NM96] Jacob Nielsen and Robert L. Mack. *Usability Inspection Methods*. John Wiley & Sons, New York, USA, 1996.
- [nod00] nodna, 3. April 2000. URL, <http://www.noDNA.com>.
- [Obe99] Carsten Oberdörster. Internet-audio und -video: Microsoft contra apple und mp3. *c't*, 10, 1999.
- [oSC00] University of Southern California. Virtuelle trainingsumgebung, 12. April 2000.

-
- [Per00] Virtual Personalities. Verbally interactive characters, 10. April 2000.
- [PG98] Ken Perlin and Athomas Goldberg. Improv: a system for scripting interactive actors in virtual worlds. In *Proceedings of the 23rd annual conference on Computer graphics*, pages 205–216. ACM, ACM, 1998.
- [PKSW96] Salvatore Parise, Sara Kiesler, Lee Sproull, and Keith Waters. My partner is a real dog: Cooperation with social agents. In *Proceedings of the ACM 1996 conference on Computer supported cooperative work*, Boston, USA, November 1996.
- [pla00] Plastic plus inc., usa: Safetycutt, Mai 2000. URL, <http://www.safetycuttblade.com/>.
- [PMG98] David Poole, Alan Mackworth, and Randy Goebel. *Computational Intelligence, A Logical Approach*. Oxford University Press, New York, USA, 1998.
- [PMP01] Ana Paiva, Isabel Machado, and Rui Prada. Heroes, villains, magicians,...: Dramatic personae in a virtual story creation environment. In *Proceedings on the International Conference on Intelligent User Interfaces*, pages 129 – 136, New York, USA, 2001. ACM, ACM Press. ISBN:1-58113-325-1.
- [Pro] www.prosieben.de.
- [Pro58] Vladimir Propp. Morphology of the folktale. *International Journal of American Linguistics*, 24(4), 1958.
- [PSP⁺92] Thomas Portele, Bernd Steffan, Richard Preuss, Walter F. Sendlmeier, and Wolfgang Hess. Hadifix - a speech synthesis system for german. In *Proceedings of the ICSLP*, pages 1227–1230, 1992.
- [PWBI98] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. In M. Ferreira, N. and Göbel, editor, *Proccedings of the 19th annual Conference of the European Association for Computer Graphics 1998*. EUROGRAPHICS, 1998.
- [RB03] Thomas Rieger and Norbert Braun. Narrative use of sign language by a virtual character for the hearing impaired. In *Proceedings of Eurographics 2003*, Granada, Spain, September 2003. Eurographics.
- [Ric98] Michael M. Richter. *Knowledge-Based Systems, Intelligence and Artificial Intelligence*. Springer Verlag, 1998.
- [Ric99] Brandon J. Rickman. The dr. k— project. In *Papers from the 1999 Fall Symposium on Narrative Intelligence, Technical Report FS-99-01*, page 120, Menlo Park, CA 94025 USA, 1999. AAAI, The AAAI Press. ISBN 1-57735-103-7.

-
- [Rie99] Roland Riempp. Intentionales bewegungslernen von bewegungs- und handlungsabläufen mit interaktivem video. Dissertation, Eberhard-Karls-Universität Tübingen, Tübingen, 1999.
- [rog00a] Guide to roguelike games, 2000. URL, <http://www.hut.fi/~eye/roguelike/>.
- [rog00b] Roguelike news, by darren hebden, 2000. URL, <http://www.skoardy.demon.co.uk/rlnews/>.
- [RP01] Volker Roth and Ulrich Pinsdorf. Semo agent plattform. www.SeMoA.org, 2001.
- [SBS96] Nitin Sawhney, David Balcom, and Ian Smith. Hypercafe: narrative and aesthetic properties of hypervideo. In *Proceedings of the the seventh ACM conference on Hypertext*, pages 1–10. ACM, 1996.
- [Sch01] Wolfgang Schneider. Software-ergonomie: Grundsätze der dialoggestaltung (iso 9241-10). http://www.sozialnetz-hessen.de/ergo-online/Software/S_Ergo-grundsaeetze.htm, 2001.
- [sea00] Searchbyvideo: Videothek als unterstützung bei der auswahl von bildungsstätten, 10. Mai 2000. URL, <http://www.searchbyvideo.com>.
- [SG00] Ulrike Spierling and Manfred Gaida. Kameraperspektiven im embassi system. Technical report, ZGDV, Darmstadt, 2000.
- [SH01] Nina Sandweg and Daniela Hermann. Analyse der technischen konzeption: Usability-anforderungen. Zwischenergebnis 5.3.1, MAP Konsortium, 2001.
- [Shn97] Ben Shneiderman. *Software Agents*, chapter Direct Manipulation Versus Agents: Paths to Predictable, Controllable, and Comprehensible Interfaces. AAAI Press / The MIT Press., Menlo Park, California, USA, 1997.
- [SM95] Chris Schmandt and Atty Mullins. Audiostreamer: Exploiting simultaneity for listening. In *Proceedings of the CHI 95*, USA, 1995. ACM.
- [SM96] Nitin Sawhney and Arthur Murphy. Espace 2: An experimental hyperaudio environment. In *Proceedings of the CHI 96*, pages 1–10. ACM, 1996.
- [SP95] Ben Shneiderman and Jenny Preece. Survival of the fittest: The evolution of multimedia user interfaces. Technical report, Department of Computer Science, Human-Computer Interaction Laboratory, and Institute for Systems Research, University of Maryland, College Park, USA, December 1995.
- [Spi00] Ulrike Spierling. Conversational integration of multimedia and multimodal interaction. In *Human Factors in Computing Systems. CHI 2000 Extended Abstracts*, pages 37–38, Den Haag, Netherlands, 2000. ACM SIGCHI, ACM.

- [Spr98] Sharon Springel. the virtual theatreimmersive participatory drama research at the centre for communications systems research, cambridge university. In *Proceedings of the sixth ACM international multimedia conference on Technologies for interactive movies*, pages 43–49. ACM, ACM, 1998.
- [SPT97] Nikitas M. Sgouros, George Papakonstantinou, and Panagiotis Tsanakas. Dynamic dramatization of multimedia story presentations. In *Proceedings of the 1997 international conference on Intelligent user interfaces*, pages 87 – 94. ACM, 1997.
- [SR00] Michael Strube and Stefan Rapp. Embassi – feinspezifikation f–pmi, pmi–d. Arbeitspapier - spezifikation pmi regelwerk, Embassi Konsortium, 2000.
- [Sti95] Lisa J. Stiefelmann. A tool to support speech and non-speech audio feedback generation in audio interfaces. In *Proceedings of the UIST’95*. ACM, 1995.
- [Sti96] Lisa J. Stiefelmann. Augmenting real world objects: A paper-based audio notebook. In *Proceedings of the Conference on Computer Human Interaction CHI ’96*. ACM, 1996.
- [Sut00] Beat Sutter. *Hyperfiktio und interaktive Narration im frühen Entwicklungsstadium zu einem Genre.*, chapter Narrationspfade in Hyperfictions; Erzählung als Weg durch den fiktiven Raum. update verlag 2000, 2000.
- [sym00] Symantec: Norton helpdesk assistant, 27. März 2000. URL, <http://www.symantec.com/sabu/nhda/index.html>.
- [TC96] Kristin R. Thorisson and Justine Cassell. Why put an agent in a body: The importance of communicative feedback in human-humanoid dialogue. In *LCC, Lifelike Computer Characters*, Utah, USA, 1996.
- [Tho96] Kristin R. Thorisson. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. PhD thesis, MIT Media Laboratory, 1996.
- [TVS] www.tv-spielfilm.de.
- [Var99] Guy Vardi. Navigation scheme for interactive movies with linear narrative. In *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots*, pages 131 – 132. ACM, 1999.
- [vHBB00] Jobst von Heintze, Christoph Bode, and Alex Butterworth. Emergent storys - eine neue herangehensweise zum erleben und zum authoring interaktiver geschichten. In Ulrike Spierling, editor, *Digital Storytelling Tagungsband*, pages 29–35, Darmstadt, Germany, 2000. Gesellschaft für Informatik und Zentrum für Graphische Datenverarbeitung, Fraunhofer IRB Verlag.
- [W3C98] World Wide Web Consortium W3C. Synchronized multimedia integration language SMIL 1.0 specification. <http://www.w3.org/TR/REC-smil/>, 1998.

-
- [Wah91] Wolfgang Wahlster. User and discourse models for multimodal communication. In J.W. Sullivan and S.W. Tyler, editors, *Intelligent User Interfaces*, pages 45 – 67, New York, USA, 1991. ACM Press.
- [Wah02] Wolfgang Wahlster. Disambiguieren durch wissensfusion: Grundprinzipien der sprachtechnologie. *KI- Künstliche Intelligenz*, (1), 2002.
- [Wal99] Jill Walker. Piecing together and tearing apart: finding the story in afternoon. In *Proceedings of the tenth ACM Conference on Hypertext and hypermedia: returning to our diverse roots*, pages 111 – 117. ACM, 1999.
- [was00] Washingtonpost: Videoarchiv der sendung "what's cooking", 10. Mai 2000. URL, <http://www.washingtonpost.com/wp-srv/style/talk/odonnell/cooking/videoarchive.htm>.
- [Web98] Dietrich Weber. *Erzählliteratur: Schriftwerk, Kustwerk, Erzählwerk*. Vandenhoeck & Ruprecht, Göttingen, Germany, 1998.
- [Wey97] Peter Weyhrauch. Guiding interactive drama. Technical Report CMU-CS-97-109, School of Computer Science, Carnegie Mellon University, Pittsburgh PA., 1997.
- [Wil97] Michael Wilson. Metaphor to personality: the role of animation in intelligent interface agents. In *Animated Interface Agents: Making them intelligent*, Nagoya, Japan, August 1997.
- [Wil98] Sue Wilcox. *Web developer.com, Guide to 3D Avatars*. John Wiley and Sons, USA, 1998.
- [Wim98] Albert K. Wimmer. *Anthology of Medieval German Literature*. 3 edition, 1998.
- [WJ95] Michael Wooldridge and Nicholas R. Jennings. Agent theories, architectures, languages: A survey. *Lecture notes in Artificial Intelligence*, Springer Verlag, 890, 1995.
- [WKL01] Manfred Weiss, Herbert Kuhlmann, and Volkmar Lotz. Multimedia workspace of the future. In *Proceedings of the Statustagung der Leitprojekte Mensch-Technik-Interaktion*. BMBF, BMBF, 2001.
- [Woo00] Steven Woodcock. Game ai. the state of the industry. <http://www.gdmag.com>, 2000.
- [WRB01] Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. Smartkom: Towards multimodal dialogues with anthropomorphic interface agents. In G. Wolf and G Klein, editors, *Proceedings of International Status Conference on Human-Computer Interaction*, pages 23 – 34, Berlin, Germany,, Oktober 2001. DLR.

- [WS98] Bob G. Witmer and Michael J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, June 1998.
- [WSS94] Janet H. Walker, Lee Sproull, and Mani Subramani. Using a human face in an interface. In *Proceedings of the CHI 94 conference companion on Human Factors in Computing Systems*, Boston, USA, 1994.
- [ZSB99] Carmen Zahn, Stephan Schwan, and Beatriz Barquero. Authoring and navigating video: Design strategies and users' needs for hyperlinks in videos - an explorative study. Report, Knowledge Communication and Knowledge Media (KCKM), Universität Tübingen, 1999.

TABELLENVERZEICHNIS

3.1	Die Aktionstypen für Image-Maps in RealSystem	52
5.1	Interaktive Funktionalitätsanforderungen für Video	106
5.2	Narrative Einsatzweise eines konversationalen Avatars	120
6.1	Abstraktionsstufen der Avatar Steuerung	131
6.2	Avatar Befehle auf Motivation Ebene	132
6.3	Avatar Befehle auf Task Ebene	133
6.4	Avatar Befehle auf Feature Ebene	133
6.5	Dialogszenario	142
6.6	Grundelemente des UIA - UIML Interface, [Ble01]	143
6.7	Grundelemente des Conversation Engine Interface	144
6.8	Konversationale Einsatzweise eines Avatars zum Video	151
7.1	EMBASSI Versuchsplan mit abhängigen und unabhängigen Variablen, [KN01]	166
7.2	Usability Prinzipien nach Höök [Hö98]	171
7.3	Beurteilung des MAP User Interface Agenten nach ISO 9241-10	176
7.4	Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, Aufgabenverteilung zwischen Mensch und Agent	177
7.5	Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, Rollenverteilung zwischen Mensch und Agent	178
7.6	Beurteilung des MAP Agentensystems nach Agenten Usability Anforderungen, sonstige Anforderungen	178
7.7	Nutzerbeurteilung der Notifikation durch einen Avatar (Voting-Angaben in Prozentzahl der Benutzer)	181
7.8	Designkriterien für die Narration von nichtlinearen Geschichten	183
7.9	Designkriterien für die Annotation von Video, [ZSB99]	185

ABBILDUNGSVERZEICHNIS

1.1	Mittelalterlicher Geschichtenerzähler (Walter von der Vogelweide) vor Publikum [Wim98]	1
1.2	Fimplakat Blade Runner	2
1.3	Online Computergame Blade Runner [Inc98]	3
1.4	Trennung der Methoden	8
2.1	Avatare der Softwarefirma NTT mit Live-Video-Gesichtern, Screenshot [Cor01]	14
2.2	Virtual Computer von Activeworlds [act00]	15
2.3	Formularbasierter Chatroom	15
2.4	Cybertown [cyb00b]	16
2.5	3D Avatar [ava00a]	17
2.6	Sicht des Benutzers in virtueller Trainingsumgebung ETOILE (etoile.tecnatom.es)	17
2.7	Moderatorin Cleo von Canal+ [cyb00a]	18
2.8	Präsentationsavatar Roxanne von RTI	22
2.9	Kiosk Assistenten, ZGDV	23
2.10	Screenshot des Office 2000, Fraunhofer IGD	24
3.1	Dramaturgie in 3 Akten nach Syd Field	30
3.2	Story Sequenz <i>Kampf mit dem Feind</i> nach Propp	34
3.3	Story Sequenz <i>Lösung einer schwierigen Aufgabe</i> nach Propp	34
3.4	Filmmontage nach Eisenstein [Eis67]	35
3.5	Eigenschaften des narrativen Designs einer Geschichte	39
3.6	Nichtlineare Erzählstruktur basierend auf Branching	40
3.7	Nichtlineare Erzählstruktur basierend auf String of Pearls	41
3.8	Nichtlineare Erzählstruktur basierend auf morphologischen Funktionen . . .	41
3.9	Synchrone und Asynchrone Interaktionswerkzeuge	64

3.10 Modell eines generativ/selektiven Storytelling-Systems, basierend auf nicht-linearem Video, mit direkt manipulativem und konversationalem Interaktionszugang	65
4.1 Menschähnliche Art der Kommunikation	70
4.2 Konversationale Interaktion in einer Mixed Reality Umgebung	73
4.3 Diskursmodell	74
4.4 Konversationsbasierte UI Struktur	75
4.5 Modalitätskanäle des Benutzers	76
4.6 Benutzer zeigt Interesse an Thema	77
4.7 Benutzer zeigt Interesse an Thema	77
4.8 Benutzer zeigt Interesse an Thema	78
4.9 Zug-Szenario zur Simulation graphisch-basierter konversationaler Interaktion	78
4.10 Auto-Szenario zur Simulation akustisch-basierter konversationaler Interaktion	79
4.11 Haus-Szenario zur Simulation graphisch-akustisch-basierter konversationaler Interaktion	79
4.12 Mimikvarianten eines virtuellen Ansprechpartners	80
4.13 Gestikvarianten eines virtuellen Ansprechpartners	81
4.14 Sensitive Zeit einer akustischen Annotation	85
4.15 Temporale akustische Annotation von akustischen Daten	85
4.16 Verschiedene Interaktionsmetaphoriken der VR, vgl. [PWBI98]	88
4.17 Temporale Darstellung eines Videohyperlink	92
4.18 Temporale Darstellung eines Videohyperlink	92
4.19 Darstellung der Dauer einer akustischen Annotation mittels Tonwiederholungen	93
4.20 Darstellung der Dauer einer akustischen Annotation mittels Reduzierung der Tondauer	93
4.21 Darstellung einer akustischen Annotation mittels Tonhöhenvariation	93
4.22 Konversation ohne virtuellen Ansprechpartner	95
5.1 Konzept Trennung applikationsspezifisches Rendering & direkt manipulative Interaktion in Kombination mit konversationaler Interaktion	102
5.2 Konzept API für nichtlineares Storytelling - Anwendungsunterstützung durch Fokussierung der Möglichkeiten	103
5.3 Skizze des Übergangsmodells der konversationalen Rederecht-Verwaltung	113

5.4	Simplifizierte Skizze der morphologischen Story Engine	116
5.5	Kameraeinstellungen	121
5.6	Avatardarstellung separat zur Informationsdarstellung	121
5.7	Avatardarstellung intergriert in Informationsdarstellung	122
6.1	EMBASSI Systemarchitektur, [HK01]	127
6.2	EMBASSI Avatar Controler und Renderer	130
6.3	Zeitliche Entwicklung der Arbeits- und Nutzungsformen an Rechnerarbeitsplätzen, [KP01]	136
6.4	MAP User Interface Agent - Kommunikation zwischen Benutzer und Basisagenten	136
6.5	<i>Working Structure</i> des MAP Projekts	137
6.6	Komponenten des User Interface Agenten und ihr Zusammenwirken, vgl. [KP01]	139
6.7	Architektur des User Interface Agenten	140
6.8	Struktur des UIA bezüglich Dialog, Konversation und abstrakten Aktoren .	141
6.9	Datenfluss im DIVA Systems	148
6.10	Applikationsspezifische Speicherung von graphischer und akustischer Annotationsinformation	148
6.11	Konversation mit menschähnlichem Avatar, angetrieben durch den Video Kontext (Screenshot des DIVA Systems)	152
6.12	Videospezifische Darstellung des Avatars	153
6.13	Videospezifische Darstellung des Avatars, Augenbraue hochgezogen	153
6.14	Videospezifische Darstellung des Avatars, Emotion <i>Veraergert</i>	154
6.15	Komponenten des DIVA Systems	155
7.1	EMBASSI Experiment - links GUI (Grundig), rechts virtueller Charakter (ZGDV), [KN01]	165
7.2	EMBASSI Experiment - Mittelwert zum Faktor <i>Interesse/Aufmerksamkeit</i> , [KN01]	167
7.3	EMBASSI Experiment - Durchschnittliche Bearbeitungszeit, [KN01]	168
7.4	EMBASSI Experiment - Durchschnittliche Anzahl Bearbeitungsschritte, [KN01]	169
7.5	EMBASSI Experiment - Anzahl natursprachliche Eingaben, [KN01]	169
7.6	MAP User Interface Agent Prototyp	177
7.7	Digitale Verteilung eines Fernsehsignals mittels eines Videoservers	183

8.1	Geist - Die Benutzerin interagiert mit einem virtuellen Avatar - einem 'geisterhaften' Bewohner des Heidelberger Schloss. Links: Benutzerin mit AR Equipement; Rechts: AR Sicht auf Katharina, einen Geist.	194
8.2	IzA, virtueller Charakter interagierend mit Benutzer und physischem Charakter	194
8.3	IzA, virtueller Charakter, physischer Charakter und Interaktionsflächen . .	195
8.4	Gebärdensprache und Mimik in einer freundlichen Ausdrucksweise	196
8.5	Gebärdensprache und Mimik in einer ärgerlichen Ausdrucksweise	196

A. ANHANG

In diesem Kapitel wird neben einer grundsätzlichen Notation der morphologischen Funktionen von Propp in Kapitel A.1 eine Erläuterung der Ansteuerung der direkten manipulativen Interaktionsmöglichkeiten in Kapitel A.2 gegeben. Weiterhin werden filmtechnische Möglichkeiten der Suspense-Erzeugung in Kapitel A.3 notiert.

A.1 *Propp, Morphologische Funktionen*

Vladimir Propp hat etwa 100 russische Volksmärchen analysiert und ein System von Funktionen notiert, welches die Volksmärchen umfassend beschreibt. Anbei wird eine Übersicht über die (wichtigsten) morphologischen Funktionen von Propp [Pro58] gegeben (ohne Berücksichtigung der zu den Funktionen bekannten Varianten). Im Anschluß werden wichtige Zusammenhänge zwischen den Funktionen erläutert.

- α Ausgangssituation
- β Abwesenheit (von Familienmitgliedern)
- γ Verbot
- δ Verstoß gegen das Verbot
- Auftritt des Feindes
- ϵ Erkundung
- ζ Antwort (der Erkundung)
- η Betrug
- θ Komplizenschaft (Aufgrund Betrugs)
- λ Heimtückische übereinkunft
- A Anschlag des Feindes
- a Mangel (morphologisch zu A äquivalent)
- B Vermittlung

C Beginn der Vergeltung

↑ Abreise

Auftritt eines Wohltäters

D Test

E Die Reaktion des Helden

F Der Held erhält einen magischen Agenten

G Transfer zu einem ausgezeichneten Ort

H Kampf mit dem Feind

I Sieg über den Feind

J Kennzeichnung des Helden mit einem Zeichen

K Das Unrecht oder der Verlust werden wieder gut gemacht

↓ Der Held kehrt nach Hause zurück

Pr Verfolgung des Helden

Rs Rettung des Helden

◦ Unerkannte Ankunft

L Ansprüche eines falschen Helden

M Schwierige Aufgabe

N Bewältigung der Aufgabe

Q Erkennung des Helden

Ex Enttarnung des falschen Helden

T Verwandlung des falschen Helden

U Bestrafung des falschen Helden oder des Feindes

W Heirat, großes Glück

Folgende Funktionen identifiziert Propp, abhängig von Charakteren, als zwingend für die jeweilige Sequenz:

- Feind: Die Funktionen A, H, Pr sind enthalten; Der Feind begeht einen Anschlag, er kämpft mit dem Helden und verfolgt ihn schließlich.

- Wohltäter: Die Funktionen D, F sind enthalten; Ein Wohltäter besitzt einen magischen Agenten und übergibt ihn.
- Helfer: Mindestens eine der Funktionen G, K, Rs, N, T sind enthalten; Ein Helfer hilft stets dem Helden.
- Gesuchte Person: Die Funktionen M, J, Ex, Q, U, W sind enthalten.
- Held: Die Funktionen C, E, W* sind enthalten; Der Held macht sich stets auf die Suche, reagiert auf die Anforderungen des Wohltäters und heiratet schließlich (d.h. erfährt großes Glück).

A.2 Explizite, temporale, intramediale Videoannotation

Die direkten manipulativen Interaktionsmöglichkeiten erfolgen über das SMIL Protokoll (Synchronized Multimedia Integration Language) [W3C98], welches vom W3C Konsortium definiert und in Kapitel 3.3.1.1 diskutiert ist. Ein Beispiel für die Videoannotationen, welche in dieser Arbeit konzipiert sind, wird im Folgenden notiert. Die nachfolgende Datei zeigt, wie SMIL zur Präsentation von expliziten, temporalen Videoannotationen, repräsentiert in Dateien mit der Endung .vhl, nutzt.

```
<smil>
  <head>
    <layout>
      <root-layout background-color="maroon" width="330"
                    height="210"/>
      <region id="videoregion1" top="5" left="5" width="320"
                    height="200" z-index="2"/>
    </layout>
  </head>
  <body>
    <par>
      <video src="Beispiel_Video.rm"      region="videoregion1"/>
      <ref src="Beispiel_VHL_1.vhl" begin="5s"/>
      <ref src="Beispiel_VHL_2.vhl" begin="12s"/>
      .
      .  weitere Videohyperlinks
      .
      <ref src="AHL_kurz.rm" begin="5s"/>
      <ref src="AHL_kurz.rm" begin="22s"/>
      .
      .  weitere akustische Annotationen
      .
      <ref src="Avatarverhalten_1.ais" begin="3s"/>
    </par>
  </body>
</smil>
```

```

        <ref src="_Avatarverhalten_2.ais" begin="38s"/>
        .
        .   weitere Avatarverhaltensequenzen
        .
    </par>
</body>
</smil>

```

Die Videoannotationsdateien (VHL-Dateien) haben ein spezielles, an den Real Media Server angepasstes Format. Eine Datei enthält sowohl Startzeiten von Hyperlinks als auch Position, Zeitpunkt, Form und Verlauf von graphischen Annotationen zu einzelnen Hyperlinks. Im Projekt DIVA sind spezielle Autorenwerkzeuge entwickelt, welche das Annotieren von Video und das Schreiben von .vhl Dateien erlauben.

A.3 *Suspense durch Filmtechniken*

Bates [BS89] gibt für die grundlegende Filmtechnik verschiedene Einstellungen vor, welche im Folgenden mitsamt ihrer narrativen Wirkung aufgelistet sind:

- *Lap dissolve* bildet eine Verbindung zwischen zwei Szenen mittels eines Fading. Konzeptioneller Link zwischen zwei verschiedenen Szenen.
- *Pan shot* [HS68] schafft einen sanften Überblick über eine Szene. Diese Technik wird benutzt, um dem Zuschauer neue Dinge mitzuteilen.
- *Strange Camera Angels*: Ungewöhnliche Kameraeinstellungen suggerieren ungewöhnliche Situationen.
- *Close-ups* geben der Szene eine Aura des Detaillierten, des sehr nahen Dabeiseins. In seiner extremen Form wird diese Technik Iris genannt.
- *Cross-cuts* verknüpfen zwei Szenen und zeigen, dass zwischen den Szenen ein Zusammenhang besteht.
- *Repetition*: Durch Wiederholung wird eine Einstellung als Schlüssel zu einem Problem dargestellt.
- *Flash shoots* oder *Quick Flashback* zeigt die Relevanz einer Szene in der aktuellen Szene.
- *Symbolic Camera Movement and Position* zeigt emotionale Einstellung der Szene.
- *Visual Rhythm* erzeugt ästhetische Konsistenz.
- *Conflicting movement* zeigt einen Konflikt an.
- *Disortion of natural rhythms* zeigt einen Konflikt ab.

- *Zoom-freeze* ergibt einen extra emphatischen Bezug des Zuschauers auf die Szene.
- *Imagery* Bildhafte Szenen geben den *Blick von Außen* auf die Szene.
- *Subjective Reality* bezeichnet Szenen, welche nicht innerhalb der objektiven Logik des Filmes liegen.
- *Non-Immersion* zeigt dem Zuschauer, dass er ein Stück Fiktion ansieht. Dies wird genutzt, um dem Zuschauer eine Distanz zu gewähren, welche ihn das Gesehene verarbeiten lässt.
- *Contrast* zwischen der Kameraeinstellung und der gezeigten Situation zeigt eine Diskrepanz zwischen der emotionalen und objektiven Bewertung einer gezeigten Situation.
- *Myoptic scenes* lassen viele kleine Szenen in der Vorstellung des Zuschauers zu einer großen Szene verschmelzen.
- *Voiceover* hat die klassische Funktion eines Erzählers.

Für die Kameraeinstellung alleine ergibt sich ebenfalls eine emotionale Wirkung, welche u.a. von [SG00] beschrieben wird. Die folgenden klassischen Kameraperspektiven werden unterschieden:

- Detail-Aufnahme: Zur besonderen Hervorhebung des gezeigten Objekts geeignet.
- Großaufnahme eines einzelnen Charakters: Hier werden die emotionalen Aspekte des Gezeigten, i.A. des Gesichtes einer Spielfigur, fokussiert. Dies führt zu einer direkten Ansprache des Publikums. Der Spannungsaufbau gestaltet sich unter dem Aspekt 'Die Spielfigur als Spiegel der Szene'. Anhand des Ausdrucks der Spielfigur kann das Publikum dessen Emotionen nachvollziehen und die Szene dadurch einfacher bewerten. Durch die Nähe der Darstellung des Schauspielers gelingt eine gewisse Verbundenheit bzw. Intimität oder Vertrautheit zum Schauspieler herzustellen, was die Immersion in die Geschichte und vor allem den Suspense in die Situation der Spiel-Figur erhöht.
- Nahaufnahme
 - Einzelner Charakter: Diese Einstellung zeigt das Gesicht und den oberen Brustbereich der Spielfigur. Dies führt zu einer Fokussierung auf die Emotionalität der Spielfigur, jedoch mit größerer Distanz zum Publikum. Dadurch ist für das Publikum eine sachlichere Wahrnehmung der Situation der Spielfigur unter Einbeziehung der emotionalen Aspekte möglich. Psychologisch bezieht sich dieses Konzept auf die zwischenmenschliche 'Individual-Distanz'.
 - Mehrere Charaktere: Wie im Falle des einzelnen Charakters besteht eine Fokussierung auf die emotionale Situation der Spielfiguren. Durch die Gruppenbildung wird der Eindruck eines vertraulichen Gespräches vermittelt. Diese Einstellung

zwingt zu einer geringen Distanz der Schauspieler untereinander, was eine gemeinsame emotionale Grundstimmung der Gruppe beschreibt. Der Zuschauer wird dabei zum Komplizen bzw. Mitwisser der Gruppe.

- Halbnahaufnahme
 - Einzelner Charakter: Die Spielfigur wird halb gezeigt (mit Kopf und Oberkörper), sie ist wahrnehmbar in der Umgebung, in der sie sich befindet. Die emotionale Situation der Spielfigur bleibt sichtbar und befindet sich im Wechselspiel mit der Umgebung, sie zeigt somit das Verhältnis der Spielfigur zur 'Welt'.
 - Mehrere Charaktere: Die Einstellung portraitiert die Schauspieler in ihrer Umgebung. Die emotionale Situation der einzelnen Spielfiguren bleibt sichtbar und kann zur Bewertung von Äußerungen in Dialogen hinzu gezogen werden. Die Grundstimmung der Spielfiguren muß nicht übereinstimmen, die Figuren stehen sich trotzdem nahe. Die Szene selbst ist geöffnet für weitere Ereignisse, z.B. hinzutretende Spielfiguren.
- 'Amerikanische Einstellung' (*knee shot*), ein und mehrere Charaktere: In dieser Einstellung können die Schauspieler schon mit vollem Gestik-Repertoire und Bewegungsradius der Arme dargestellt werden, jedoch sind sie körperlich noch nicht vollständig im Bild. Sie eignet sich damit gut, um Aktionen von Spielfiguren in ihrer direkten Umgebung darzustellen. Die Emotionale Stimmung der Szene ergibt sich für das Publikum zu gleichen Teilen aus den Aktionen und der Mimik der Spielfiguren - die Mimik ist noch gut erkennbar.
- Halbtotale, ein und mehrere Charaktere: Die Spielfiguren werden in ihrer vollen Größe und mit der ganzen Palette der körperlichen Ausdrucksmöglichkeiten gezeigt. Diese Einstellung eignet sich zur Darstellung von Aktionen, die Emotionalität der Spielfiguren kann für das Publikum durch deren Handlungen erschlossen werden, die Mimik tritt als emotionaler Schlüssel zur Szene zurück.

B. SCHRIFTENVERZEICHNIS

B.1 Eigene Publikationen

- Braun, Norbert, :
Storytelling & conversation to improve the fun factor in software applications
In: Mark A. Blythe, Andrew F. Monk, Kees Overbeeke, and Peter C. Wright (ed.):
Funology, From Usability to Enjoyment, Chapter 19, Kluwer Academic Publishers,
Dordrecht, ISBN 1-4020-1252-7, April 2003.
- Braun, Norbert; Rieger, Thomas:
Expressiveness Generation for Virtual Characters in Games
In: Level Up Gamesconference, Digital Games Research Conference 2003, 4-6 November 2003 University of Utrecht, The Netherlands.
- Oliver Schneider, Braun, Norbert:
Content Presentation in Augmented Spaces by the Narration of Interactive Scenes
In: in: First Research Workshop on Augmented Virtual Reality AVIR03, University of Geneva, Switzerland, September 17-18, 2003.
- Braun, Norbert
Enter the Digital Wonderland using the Desktop Metaphor? About the Advantage of Using Narrative Computer Human Interfaces
in: Mensch & Computer 2003, 3. Fachübergreifende Konferenz, Interaktion in Bewegung, Stuttgart, Germany, September, 07 - 10, 2003.
- Rieger, Thomas; Braun, Norbert:
Narrative Use of Sign Language by a Virtual Character for the Hearing Impaired
in: Conference Proceedings of Eurographics 2003, Granada (Spain), September, 1. -6. 2003, 2003.
- Braun, Norbert; Rieger, Thomas :
Expressiveness Generation for Virtual Characters, based on the Principles of Suspense Progression and Narrative Conflict
in: Conference Proceedings Conference Proceedings of the International Workshop on Mobile Computing, IMC, Rostock, Germany, June, 17-18, 2003.
- Goebel, Stefan; Braun, Norbert; Spierling, Ulrike; Dechau, Johanna; Diener, Holger:
Proceedings Preamble TIDSE 03

- In: Goebel, Stefan; Braun, Norbert; Spierling, Ulrike; Dechau, Johanna; Diener, Holger; Technologies for Interactive Digital Storytelling and Entertainment TIDSE 03. Conference Proceedings, Darmstadt, Germany, 2003
- Hellenschmidt, Michael; Braun, Norbert:
Developing Context- and User Groups Sensitive Learning Scenarios with XML Configuration
In: HCII International 2003, 10th International Conference on Human - Computer Interaction, Crete, Greece, 2003.
 - Braun, Norbert; Schneider, Oliver:
Suspenseful User Experiences in Collaborative Virtual Spaces, Enabled by Interactive Narration
In: HCII International 2003, 10th International Conference on Human - Computer Interaction, Crete, Greece, 2003.
 - Braun, Norbert:
Storytelling in Collaborative Augmented Reality Environments
In: Skala, Vaclav; EUROGRAPHICS, Computer Graphics Society and IFIP Working Group 5.10 on Computer Graphics and Virtual Worlds: WSCG 2003. Conference Proceedings. Plzen : University of West Bohemia, 2003
 - Schneider, Oliver; Braun, Norbert:
Storylining Suspense: An Authoring Environment for Structuring Non-linear Interactive Narratives
In: Skala, Vaclav; EUROGRAPHICS, Computer Graphics Society and IFIP Working Group 5.10 on Computer Graphics and Virtual Worlds: WSCG 2003. Conference Proceedings. Plzen : University of West Bohemia, 2003
 - Braun, Norbert; Dechau, Johanna; Schneider, Oliver; Goebel, Stefan:
The Story as Contextual Basis of the Collaborative Group Experience
In: ACM Collaborative Virtual Environments 2002, Workshop on Storytelling in Collaborative Virtual Environments, Bonn, Germany, 2002
 - Braun, Norbert, Schneider, Oliver, Habinger, Gregor:
Literary Analytical Discussion of Digital Storytelling and Its Relation to Automated Narration
In: HCI Europe '2002, Workshop Understanding User Experience: Literary Analysis meets HCI; London, UK , 2002
 - Braun, Norbert:
Automated Narration - the Path to Interactive Storytelling
In: Workshop on Narrative and Interactive Learning Environments, Edinburgh, Scotland, 2002
 - Braun, Norbert:
Programmierte Narration und Konversation, Eingesetzt zum Interaktiven Erzählen

von Geschichten.

In: Mensch und Computer 2002 - Vom interaktiven Werkzeug zu kooperativen Arbeits- und Lernwelten, Hamburg, Germany, 2002

- Braun, Norbert:
Narrative Semiotics for Computer Games.
In: Challenge of Computer Games, Lodz, Poland, 2002
- Braun, Norbert:
Menschähnliche Konversation - eine diegetische Narrationskomponente für Delegation und Assistenz.
In: VDI/VDE - GMA Fachtagung Useware 2002, Darmstadt, Germany, 2002
- Braun, Norbert:
Storytelling & Conversation to Improve the Fun Factor in Software Applications.
In: CHI 2002 Conference on Human Factors in Computing Systems, Workshop Funologie, Minneapolis, Minnesota, SIGCHI, ACM, USA, 2002.
- Braun, Norbert:
Symbolic Conversation Modeling Used as Abstract Part of the User Interface.
In: Skala, Vaclav; EUROGRAPHICS, Computer Graphics Society and IFIP Working Group 5.10 on Computer Graphics and Virtual Worlds: WSCG 2002. Conference Proceedings. Plzen : University of West Bohemia, 2002
- Spierling, Ulrike; Grasbon, Dieter; Braun, Norbert; Iurgel, Ido
Setting the Scene: Playing Digital Director in Interactive Story Telling and Creation
In: Encarnação, José L. (Hrsg.): Computers and Graphics, An International Journal of Systems & Applications in Computer Graphics. Darmstadt, Germany, 2002
- Braun, Norbert:
Modeling of Conversational User Interface
Tutorial at International Conference on Text Speech and Dialogue, TSD2001, Faculty of Informatics, Masaryk University, Brno, Faculty of Applied Sciences, University of West Bohemia, Plzeň (Pilsen), Czech Republic, 2001
- Grasbon, Dieter; Braun, Norbert:
A Morphological Approach to Interactive Storytelling
In: Fleischmann (Hrsg.) u.a.: Proceedings of the Conference on artistic, cultural and scientific aspects of experimental media spaces. Bonn, Germany, 2001
- Braun, Norbert; Rieger, Thomas:
Group Conversation within a Internet TV Community through a Lifelike Avatar
In: Proceedings of the International Conference on Media Futures, Italien, 2001
- Braun, Norbert; Finke, Matthias; Rieger, Thomas:
Community TV: An Approach to Interaction for Groups and Single Users on Internet Video.

In: Kluev, V.V. (Hrsg.) u.a.; World Scientific and Engineering Society (WSES): Proceedings of WSES Conferences 2001. CD-ROM : SSIP-MIV-SIM-RODLICS. 2001

- Braun, Norbert; Schneider, Oliver:
Conversation Modeling as an Abstract User Interface Component.
In: Bauknecht, Kurt (Hrsg.) u.a.; Gesellschaft für Informatik (GI) u.a.: Informatik 2001. Tagungsband der GI / OCG Jahrestagung : Wirtschaft und Wissenschaft in der Network Economy - Visionen und Wirklichkeit. Wien : Österreichische Computer Gesellschaft, 2001, S.1223-1230
- Braun, Norbert:
Interaction Approach for Digital Video Based Storytelling.
In: Skala, Vaclav; EUROGRAPHICS, Computer Graphics Society and IFIP Working Group 5.10 on Computer Graphics and Virtual Worlds: WSCG 2001. Conference Proceedings Vol. 2. Plzen : University of West Bohemia, 2001, S. 367-374
- Müller, Wolfgang; Alexa, Marc; Rieger, Thomas; Braun, Norbert:
Ein flexibles Präsentationssystem für animierte User-Interface-Agenten und Avatare.
In: Spierling, Ulrike (Hrsg.): Digital Storytelling - Tagungsband. Stuttgart : Fraunhofer IRB Verlag, 2000, S. 163-175 (Computer Graphik Edition 2).
- Braun, Norbert; Finke, Matthias:
Interaction of Video on Demand Systems with Human Like Avatars and Hypermedia.
In: Scholten, Hans (Hrsg.) u.a.: Interactive Distributed Multimedia Systems and Telecommunication Services. Proceedings 2000. International Workshop. Berlin; Heidelberg : Springer, 2000, S. 172-186 (Lecture Notes in Computer Science 1905).
- Braun, Norbert; Gaida, Manfred; Lohde, Ulrike; Spierling, Ulrike:
Scenario-based Design - Geschichten als Hilfsmittel beim Erstellen von Software.
In: Spierling, Ulrike (Hrsg.): Digital Storytelling - Tagungsband. Stuttgart : Fraunhofer IRB Verlag, 2000, S. 13-28 (Computer Graphik Edition 2).
- Weber, Claudia; Braun, Norbert; Spierling, Ulrike:
Neue Entwicklungsmethoden für Digital Storytelling - Ein Vergleich der etablierten Bereiche der Filmproduktion und der Software-Entwicklung und ihre Relevanz für die moderne Multimedia-Anwendungsentwicklung In: ZGDV e.V., Reportnr.: 99s001-ZGDV, Darmstadt, 1999
- Braun, Norbert:
Conversational Video Interaction. In: Bra, Paul de (Hrsg.) u.a.; Association for the Advancement of Computing in Education: WebNet 99 - World Conference of the WWW and Internet. Proceedings. Charlottesville, VA, USA : Association for the Advancement of Computing in Education, AACE, 1999, S.1200 - 1201
- Braun, Norbert; Dörner, Ralf:
Temporal Hypermedia for Multimedia Applications in the World Wide Web.

- In: Veruna, Brijesh u.a.; Institute of Electrical and Electronics Engineers: International Conference on Computational Intelligence and Multimedia Applications 1999. Los Alamitos, Calif. : IEEE Computer Society, 1999, S. 413-417
- Braun, Norbert; Dörner, Ralf; Soetebier, Ingo:
A VRML and Java-Based Interface for Retrieving VRML Content in Object-Oriented Databases.
In: Bra, Paul de (Hrsg.) u.a.; Association for the Advancement of Computing in Education: WebNet 99 - World Conference of the WWW and Internet. Proceedings. Charlottesville, VA, USA : Association for the Advancement of Computing in Education, AACE, 1999, S. 987-992
 - Soetebier, Ingo; Dörner, Ralf; Braun, Norbert:
Seamless Integration of Databases in VR for Constructing Virtual Environments.
In: EUROGRAPHICS, Computer Graphics Forum 18 (1999), 3, S. C-331 - C-337
 - Braun, Norbert; Dörner, Ralf:
Sonic Hyperlinks: Hypermedia Methodologies Applied to Audio for WWW-Based Teaching Applications.
In: Maurer, Hermann u.a.; Association for the Advancement of Computing in Education: WebNet 98 - World Conference of the WWW, Internet & Intranet. Proceedings. Bd. 1. Charlottesville, VA, USA : Association for the Advancement of Computing in Education, AACE, 1998
 - Braun, Norbert; Dörner, Ralf: Using Sonic Hyperlinks in Web-TV. In: Department of Computing Science, University of Glasgow: Auditory Display, ICAD '98. Papers. Glasgow, 1998
 - Braun, Norbert; Blechschmitt Eric: Studie Business TV In: Fraunhofer IGD, Reportnr.: 97v001-FIGD, Darmstadt, 1997
 - Braun, Norbert; Englert, Gabriele; Höppner, Silke; Wiener, Andreas; Yildirim, Feramuz: MM-Datenbanken, Archive und Informationsdienste. Abschlussbericht In: In: Fraunhofer IGD, Reportnr.: 96i009-FIGD, Darmstadt, 1996

B.2 Betreute Diplomarbeiten

- I. Soetebier: Realisierung und Evaluierung eines internetbasierten 3D - Datenbankinterfaces für ein verteiltes 3D - Animationssystem. Fachhochschule Darmstadt, Darmstadt, 1998.
- J. Laserer: Internet Video-Mate: Konversationales Interface zur Präsentation interaktiver Videos. Fachhochschule Hagenberg, Hagenberg, Österreich, 2000.
- A. Graham: Client/Server Videosystem mit integrierter Hypermedia- und Konversationsinteraktion. Fachhochschule Darmstadt, Darmstadt, 2000.

- D. Grasbon: Konzeption und prototypische Implementation einer Storyengine: Dynamisch-reaktives System zum Erzählen nichtlinear-interaktiver Geschichten bei größtmöglicher Spannung, gedanklicher Immersion, Identifikation und Motivation des Spielers. Technische Universität Darmstadt (TUD), Fachbereich Informatik (FB 20), Fachgebiet Graphisch-Interaktive Systeme (GRIS), Darmstadt, 2001.

C. LEBENSLAUF

Name:	Norbert Braun	
Geboren am:	18.07.66	
Geboren in:	Odernheim (Glan)	
Ausbildung:	1972 - 1985	Schulbildung mit Abschluß Abitur, Gymnasium <i>Pädagogisches Erziehungsinstitut Alzey</i>
	1986 - 1994	Studium der Informatik mit Nebenfach Wirtschaftswissenschaften an der Universität Kaiserslautern, Abschluß Diplom Informatik
Berufspraxis	1990 - 1994	Tätigkeit als wissenschaftliche Hilfskraft der Universität Kaiserslautern
	1994 - 1995	CGTec/Sema, Abteilung Forschung und Entwicklung
	1996 - 1998	Wissenschaftlicher Mitarbeiter, Technische Universität Darmstadt (TUD), Fachbereich Informatik (FB 20), Fachgebiet Graphisch-Interaktive Systeme (GRIS), Darmstadt
	1998 - 2002	Wissenschaftlicher Mitarbeiter am Zentrum für Graphische Datenverarbeitung, Abteilung Digital Storytelling, Darmstadt
	seit 2003	Wissenschaftlicher Mitarbeiter, Technische Universität Darmstadt (TUD), Fachbereich Informatik (FB 20), Fachgebiet Graphisch-Interaktive Systeme (GRIS), Darmstadt